2020-01-28

Baby bumps in the road: The impact of parenthood on job performance, human capital, and career advancement

Healy, Olivia; Heisell, Jennifer

SSRN

Baby bumps in the road: The impact of parenthood on job performance, human capital, and career advancement*

Olivia Healy
Cornell University

Jennifer Heissel[a]
Naval Postgraduate School

January 28, 2022

Abstract

A robust literature establishes that women experience lasting penalties to their labor market earnings after having a child, but men do not. This paper explores possible factors behind earnings penalties. We evaluate the effect of parenthood on men's and women's job performance, human capital accumulation, and career advancement using unique personnel data on U.S. Marines. To address selection of workers into parenthood, we estimate event study models around the first birth of a child. We include a comparison group of nonparents assigned "placebo births" based on key predictors of parenthood and outcomes, which we argue best approximates the counterfactual. Outcomes include standardized measures of job performance, such as physical fitness test scores, supervisor ratings of job proficiency, and firm-specific task scores (rifle and pistol marksmanship). We also study time spent in firm-specific training, formal years of education, and promotions.

Among U.S. Marines who return to work with similar hours after having a child, we find declines in mothers' job performance and months of firm-specific training in the two years following a birth. Consistent with these findings, women's promotion trajectories slow in response to childbirth. Men have slightly lower physical performance in the first year postbirth, but other performance measures, human capital development, and promotion trajectories are largely unaffected. Mediation analyses suggest delays in promotions stem from mothers missing key job performance tests due to pregnancy and postpartum waivers. Last, we show longer paid maternity leave due to an unexpected leave extension does not predict better or worse job-related outcomes.

JEL Classification: J24, J16, J18, J45
Keywords: Parenthood, Child Penalty, Gender Wage Gap, Promotion, Parental Leave
Word Count: 10,929

[a]*Corresponding Author.* 1 University Cir., Monterey, CA 93908; jaheisse@nps.edu

*The authors are grateful to Maj. Tamara Cordero, Capt. Amanda Henegar, Capt. Michael Larson, Maj. Laura Laurito, Capt. Julie Schumacher, Maj. Genevieve Studer, Capt. Ansley White and all Naval Postgraduate School Marines who provided insight into Marine processes and procedures. We also thank Jonathan Guryan, Matthew Notowodigdo, Diane Schanzenbach, and seminar participants at the American Economic Association, Society of Labor Economists, Midwestern Economic Association, Eastern Economic Association, Western Economic Association, Association for Public Policy Analysis and Management, and Stanford University for helpful comments. We are grateful for funding from the Robert Wood Johnson Foundation's Evidence for Action grant program (77124); the Institute of Education Sciences' Multidisciplinary Program in Education Sciences (Award R305B140042); and Northwestern University's Graduate Research Grant Program. Any errors are those of the authors. The views expressed in this paper do not reflect the views of the U.S. Department of Defense or the U.S. Navy. Note that the data from the Department of Defense are not available for sharing with the public.

I Introduction

After having a first child, women experience large and persistent declines in earnings but men do not. This statistical fact, termed the "child penalty," is well-established and holds across countries and contexts (Aguilar-Gomez et al., 2019; Andresen and Nix, 2020; Angelov et al., 2016; Barth et al., 2017; Bertrand et al., 2010; Kleven et al., 2019a,b). Maternal child penalties also explain much of the overall gender earnings gap between men and women in higher-income countries (Cortes and Pan, 2020; Kleven et al., 2020, 2019b). Prior research shows that three main factors drive gendered child penalties following birth: (1) exit rates from the labor market, (2) hours worked conditional on employment, and (3) wages (Kleven et al., 2019b). Yet, the mechanisms driving these gendered employment responses are less understood. Few data sources track men's and women's work outcomes over time and at the level of detail needed to investigate why women's labor market participation responds differently to parenthood than men's. Better evidence on the mechanisms behind the gender-based responses to childbirth will allow policymakers to develop strategies to address child penalties.

We use administrative records on service members in the U.S. Marine Corps to overcome some of the limitations of prior research. Our primary advantage is access to detailed, consistently measured, longitudinal data on *precursors* to worker output and productivity. In contrast to standard labor market measures (e.g., employment, hours worked, and wages),[1] we examine high-frequency measures of individuals' job performance, job-specific and general human capital accumulation, and career advancement. A second strength is that workers largely stay on the job working similar hours after becoming parents due to multiyear contract requirements in the Marines. As such, our focus on this single employer helps hold two key drivers of the child penalty constant, limiting selection out of the workforce and minimizing reductions in hours worked in response to parenthood. In turn, our empirical analysis sheds light on the following question: if men and women were to stay on the job working similar hours after childbirth, would a child penalty in earnings emerge, and why?

Two possibilities in the immediate postbirth period could explain why a child penalty might still

[1] A longstanding literature on the impact of fertility on parents' employment outcomes has focused on hours, employment, and wages; these are often the only measures available (Agüero and Marks, 2011; Angrist and Evans, 1998; Bronars and Grogger, 1994; Cáceres-Delpiano, 2006; Cools et al., 2017; Cruces and Galiani, 2007; Jacobsen et al., 1999).

emerge among parents who remain on the job working similar hours. First, mothers may struggle more than fathers to keep up their performance after having a child, whether due to greater child care responsibilities, the unique biological impact of birth, or other constraints. If mothers struggle on the job after returning to work, they may later exit the labor market or reduce hours worked down the road. Then, even among those who stay employed, medium- or long-term changes in labor market participation could give rise to lasting maternal child penalties. Second, possibility is that if women struggle to perform and advance their job skills after having children, their promotion rates may slow. Delays in career advancement due to childbearing could damper mothers' wage growth relative to fathers'. If mothers' career progression does not catch up, child wage penalties could persist in the long run.

We explore empirical evidence behind these two possibilities. We analyze the impacts of a first birth on job performance and human capital development, which jointly determine promotion decisions in the Marines. We then estimate direct impacts on promotions. Our data allow a fine-grained assessment of the month-by-month consequences of parenthood in the immediate two years following a birth. Measures of job performance include scores on Marine-specific assessments of physical fitness performance and marksmanship skills, as well as supervisor ratings of overall job performance. Human capital outcomes include months spent in job-specific training and the accumulation of years of formal education. Career advancement tracks the number of promotions Marines receive over time. We explore changes in these outcomes among a sample of parents who remain employed for at least two years after having a child.

Our empirical strategy uses the precise month of childbirth as an exogenous shock to parents' work performance outcomes. We use data on Marines who do not have a child during the study window to account for secular time trends in outcomes, and in our preferred estimate we use a matching strategy that assigns nonparents to "placebo births." We exact match parents and nonparents on tenure (in months) in the Marine Corps, job rank, being active duty (full-time) or a reservist (part-time), and calendar year. We further restrict the matches based on LASSO-selected observable characteristics that best predict the likelihood of a first birth. Using this strategy, we compare outcomes between parents and nonparents before and after a birth (or placebo birth). The placebo birth matching strategy better corrects for observable differences between parents and nonparents that might affect parents' postbirth outcome trajectories in the absence of a having a child, compared to a traditional two-way fixed effects model. The matching strategy allows us to study cumulative changes in outcomes (e.g., total months spent in training during pregnancy and posbirth), given that we can align parents and nonparents by event time and measure changes in the outcome relative to the event/placebo event. The matching strategy also allows us to explore subgroup differences based on prepregnancy characteristics.[2] We estimate all effects separately for women and men because we want to explore whether parenthood also affects men along the margins we study. If men are affected, fathers would not be a good comparison point for mothers.

We find persistent effects of parenthood on work performance, firm-specific human capital accumulation, and promotion trajectories for mothers. Mothers' physical fitness performance, job performance, and marksmanship scores decline immediately postbirth, as soon as they are observed again (after any testing exemptions around pregnancy/birth are lifted). Mothers' job performance and marksmanship scores recover by their child's second birthday, but their physical fitness scores remain 0.2 standard deviations lower than prepregnancy. Mothers also accumulate fewer months of training relative to nonmothers, with gaps beginning during pregnancy and continuing through two years postbirth. Consistent with these patterns, mothers' promotion trajectories slow as a result of having a child, with gaps widening, rather than narrowing, during the two years postbirth.

We observe minimal impacts of parenthood on fathers' outcomes. The birth of a child leads to small, short-lived declines in physical performance, job performance, and training one month postbirth. Fathers' performance on all four measures recovers by the child's second birthday. We do not observe a consistent patterns of effects on fathers' years of formal education or accumulated promotions postbirth.

Taken together, our results suggest that the immediate period after having a child may be uniquely challenging for women who remain employed. Yet, it is unclear whether declines in job performance and human capital accumulation that we document give rise to observed promotion delays for mothers in our setting. To explore this possibility, we use a back-of-the-envelope approach and control for endogenous outcomes of childbirth (changes in job performance and human capital accumulation) when predicting

[2] In an event study approach, nonparents help to estimate secular time trends in the outcome (i.e., time fixed effects), but they are never assigned event time given that they do not experience the event. As such, it is not possible to pinpoint pre/postbirth periods for nonparents and use their outcomes as a counterfactual. When we assign placebo births, it creates a clear pre/post timeframe for control cases.

the impact of birth on promotions. We find declines in mothers' physical fitness performance scores partly explain the impact of birth on promotions. However, the biggest factor driving mothers' promotion delays is the number of *missing* physical fitness performance assessments mothers accumulate due to pregnancy- and birth-related test waivers. Mothers are waived from taking fitness assessments during pregnancy and at least 6 months postbirth, but other individuals may miss these tests due to idiosyncratic mishaps like sprained ankles. Performance on fitness assessments is an explicit component of the promotion decision. We find missing physical fitness assessments harms promotion rates for all Marines, and once we control for missing these tests, mothers' promotion trajectories are indistinguishable from matched nonmothers.

Last, we explore whether variation in maternity leave length predicts larger or smaller impacts of childbirth on mothers' performance, skill accumulation, or promotion trajectories. To address possible selection into leave length or parenthood due to leave length, we focus on differences across groups of women who were already pregnant (or who had already given birth) when maternity leave length was unexpectedly extended. We compare three distinct policies: 6 weeks of leave (baseline), 6 weeks plus 12 flexible weeks of leave after returning to work, and 18 weeks weeks of leave. We do not find statistically significant differences in the magnitude of parenthood impacts across the three policies. Longer leave may allow mothers to invest in non-work domains, including their own or their children's health and well-being.[3] We take it as good news that longer leave does not hamper job performance, skill development, or promotion outcomes for mothers in our setting.

Most similar to our paper, three studies focus on gender-based differences in worker output across the transition to parenthood. Azmat and Ferrer (2017) show female lawyers with young children are less productive compared to male lawyers with young children in terms of annual hours billed, and Kim and Moser (2020) find female scientists in the 1950s patented less during their childbearing years, relative to fathers and other women. Gallen (2018) explores how firm output in Denmark varies by the gender and

[3]For example, Balser et al. (2020) find maternity leave extensions from 6 to 12 weeks reduce postpartum depression diagnoses among mothers in the Army and Air Force. In U.S. and international contexts, access to paid parental leave improves mothers' physical and mental health (Bullinger, 2019; Butikofer et al., 2018), duration of breastfeeding (Pac et al., 2019), infant health (Bullinger, 2019), and time spent with children after return to work (Bailey et al., 2019; Trajkovski et al., 2019).

parenthood status of employees in private firms. She finds mothers are substantially less productive—according to a firm production function model—than other workers, particularly during childbearing years. Our paper focuses on a precursor to work output: work performance and skill accumulation.

A key contribution of our paper is the ability to trace the dynamic responses of men and women month by month in the two years following a first birth. Prior papers that estimate child penalties use yearly measures of earnings or income and therefore cannot detect initial, within-year impacts (Andresen and Nix, 2020; Angelov et al., 2016; Barth et al., 2017; Bertrand et al., 2010; Kleven et al., 2019a,b). We add to the existing literature by providing evidence that parenthood begins to impact women's performance *within* the first year after birth. Results support the notion that immediate postbirth declines in mothers' ability to perform at work and advance their job skills may lead mothers to exit the labor market, reduce hours worked, or garner lower wages by the time these outcomes are measured one or more years postbirth in other papers. Our finding that childbirth slows mothers' but not fathers' promotion trajectories suggests that even absent changes in employment and hours worked, having a child can create gender-based earnings penalties through its impact on mothers' career advancement.

Given the size and role of the U.S. military, our findings on gender gaps in Marine parents' work outcomes are important in their own right. Gender-issues are prevalent in the U.S. military, and women who choose to serve in this male-dominated setting may be, on average, more committed to military careers than men for whom the decision to serve is more common. Declines in work outcomes for female service members, then, may be especially noteworthy. Efforts to increase female presence in the military, especially at higher ranks, require supporting mothers' physical recovery and career advancement. Beyond the importance of our findings for the military setting, our results also provide evidence for workers not currently represented in the literature, specifically those in physically demanding occupations. About 45% of jobs in the civilian labor market require at least medium physical strength, defined as work that involves frequent lifting or carrying of objects weighing up to 25 pounds (Bureau of Labor Statistics, 2017). Baseline physical health is a relevant form of human capital, emphasizing the importance of our findings on job performance and work outcomes in the Marine Corps setting.

II Institutional Background

The DoD is the world's largest employer, with a total of 1.3 million active-duty service members. We focus on the Marine Corps, where administrative records on job performance are readily available. Marines are an immediate response force, ready to deploy quickly to support combat missions on sea or land. Marines make up nearly 15% of active-duty forces. Our findings may be particularly relevant for male-dominated industries or organizations, as 92% of Marines are male (Department of Defense, 2018).

Individuals begin in the Marine Corps either as a junior enlisted, akin to an entry-level civilian worker; or as an officer, akin to a civilian manager.[4] There are over 35 career fields in the Marines, and each has dozens of specializations, referred to as Military Occupational Specialties. Some career fields are specific to the military (e.g., infantry); others are also present in the civilian labor market, including food services, financial management, military police and corrections, and legal services.

A service member's occupational specialty and their assigned unit determines their day-to-day work environment. Our analytic sample predominantly consists of active-duty Marines who work full time, Monday through Friday. For these individuals, the workday typically begins with early morning physical training, followed by work assignments through the evening. Most service members live and work on or near a military base and are stationed in the U.S. (83% of service members; Department of Defense 2018), though some are stationed abroad. Our sample also includes reservists, whose service requires (1) participating in training drills one weekend per month and (2) attending a two-week program each year. Reservists typically have civilian careers or are enrolled in higher education while they fulfill part-time Marine Corps reserve duties. Reservists can be called upon for active-duty deployment in times of war or national emergency, at which point they are active reserves and work for the Marines full-time. We include active reservists as active-duty.

Marines sign a legally binding contract that outlines their required length of service. Initial enlisted contracts typically require 4 years of active-duty service, while the officer commitment is typically 3

[4]Enlisted service members must have a high school degree and be between 18 and 29 years old when they begin. Officers must have at least a bachelor's degree upon entry. Roughly 89% of Marine Corps service members are enlisted, and most are of prime childbearing age. Among Marines, 81% are under 30 years old, and more than one quarter are parents (Department of Defense, 2018).

years. These contracts limit the extent to which Marines can exit the labor force after they have a baby.

Effective job performance in the Marines requires both mental and physical acuity. The Marine Corps uses a standardized set of measures to evaluate performance among both active-duty and reserve Marines. Performance measures include physical performance, supervisor-scored job proficiency, and marksmanship tests. Scores on these measures are used to determine promotions, with different weights placed on each depending on the given promotion juncture.[5] Marines also go through long-term training exercises to develop job-specific skills (e.g., pilot training); training can last from a few weeks to several months. Many Marines also pursue educational degrees while on the job, either paid for as part of their job (e.g., being given orders to obtain a master's degree full-time) or paid for privately (e.g., an enlisted Marine pursuing a bachelor's degree outside of work).[6] Training and education also contribute to promotion decisions. Based on these clearly identified measures, Marines can determine what they need to advance and, therefore, have especially strong incentives to perform well on measured assessments.

The DoD provides a number of family-friendly benefits, including fully paid parental leave. In a supplementary analysis, we focus on policy changes to the length of paid leave for primary caregivers (most often women) and refer to this as variation in maternity leave length.[7]

III Data

We draw data from the Marine Total Force Data Warehouse and obtain records on all active-duty and reserve Marines who served at any point during January 2010 through December 2019. Our data include basic descriptive information on service member characteristics (age, gender, race/ethnicity, education status, and AFQT/GCT scores—which measure aptitude and intelligence), dependent characteristics for spouses and children (exact date of birth, gender, and whether a spouse is in the military), and job

[5]Promotions at lower ranks, E1 through E3, are relatively automatic after a given number of months in service and months in rank. A composite score based on performance metrics determines promotion from the rank of E4 to E5, conditional on meeting requirements for minimum time in service and minimum time in the current job level to be eligible for promotion. For promotion at ranks above E5, the same performance metrics, along with supervisor ratings, are reviewed by an evaluation board to determine promotion. As such, Marine Corps promotions are similar to civilian promotions. Both are based on work performance, but the Marine Corps promotion system is arguably more systematic.

[6]Marines who obtain education as part of their job generally add time to their commitment as part of a "pay-back tour."

[7]Paternity leave changed from 10 to 14 days during this time. We do not examine paternity leave in this paper.

characteristics of the service member (job type, professional rank, time in service, and unit location).

Our first set of outcomes consists of the three primary measures of job performance used for promotion and retention decisions. The first of these outcomes, physical performance, captures Marines' ability to perform physically demanding work tasks, as measured by standardized fitness tests. Marines take two tests per year: the physical fitness test (timed running, crunches, and pull-ups/push-ups) in the first half of the year and a combat fitness test (timed running, a combat-related obstacle course, and upper-body strength measured by ammunition can lifts) in the second half of the year. Scores are awarded on a 300-point scale, which is adjusted for age and gender such that women do not need to do as many pull-ups as men, and older service members do not need to run as fast as younger ones to achieve the same score. We standardize points-based physical fitness scores by year, gender, and test type. We combine the Z-scores for the two tests into one measure, generally observed twice per year per Marine. Due to the physically demanding nature of the assessment, women are not required to take the test when pregnant, and they are exempt from tests 6 months postbirth. We resume measurement of the physical performance outcome for women at 8 months postbirth due to concerns that commanding officers may allow some women whose test dates fall 7 months after birth to skip the test during that assessment round.

The second job performance outcome we measure consists of supervisor evaluations of Marines' work performance. Supervisors regularly assess Marines' using one of two rating scales, depending on the Marine's rank. Junior enlisted receive what are called proficiency and conduct marks ("ProCons"), and senior enlisted and officers receive ratings called Fitness Reports ("FitReps"). Both assessments require supervisors to assess a Marine's performance across a range of professional domains, such as technical knowledge, effectiveness, and communication skills.[8] We standardize supervisor ratings by year, gender, and assessment type. We then combine the Z-scores into one outcome we call job performance. We generally observe ratings at least twice per year among junior enlisted, once in the first half and once in the second half. For senior enlisted and officers, we observe supervisor ratings a minimum of once per year. If a Marine is transferred, discharged, or promoted, or if their supervisor changes, they

[8]The full set of domains includes mission accomplishment (job-specific aptitude, competence, technical knowledge, and practical skills), character (courage, effectiveness under stress, and initiative), leadership (setting the example, communication skills), and intellect and wisdom (professional military education, decision-making ability, and judgement), among others.

receive additional performance ratings.[9] We are missing supervisor ratings for junior enlisted who left the Marines before October 2017. However, we observe the full history of performance ratings (including prior to October 2017) for any Marine who was active as of October 2017. The subjective nature of supervisor ratings means we cannot distinguish true changes in job performance from supervisors' *perceptions* of changes in performance using this measure.

Our third job performance outcome captures rifle and pistol marksmanship assessment scores. Marksmanship assessments evaluate Marines on their target shooting skills and award points according to performance. While this measure is clearly a firm-specific measure of job proficiency, strong marksmanship performance is partially cognitive and requires practice and concentration. We standardize marksmanship scores by year, gender, and weapon (rifle/pistol). The marksmanship assessment is required once a year at junior levels and becomes optional for more senior Marines. Those who perform at the highest level (rated "expert") are exempt from re-testing the following year, making outcome data on this measure more sparse. Like physical performance tests, the marksmanship test requirement is waived for pregnant mothers and not assessed when mothers are on maternity leave. We resume measurement of marksmanship scores for mothers starting 5 months postbirth, given that some mothers in our study window could have been on leave for 18 weeks.

Figure I displays counts of physical performance, job performance ratings, and marksmanship scores relative to birth for mothers; the figure shows a drop in the frequency of observations around the birth event for physical performance and marksmanship scores, corresponding to periods during which women are generally exempt from tests. We exclude unshaded observations of women's physical performance and marksmanship scores from our analyses.

Our second set of outcomes captures human capital development. To measure firm-specific human capital accumulation, we observe whether or not a Marine is in training on a monthly basis. We create a cumulative count of months of training observed for each Marine relative to 10 months before they have a child. A value of 1 on the variable indicates 1 additional month of training achieved since the time point before pregnancy. To measure general human capital, we track increases in Marines' formal education

[9]Marines are relocated every few years, as are their designated supervisors. Decisions on relocation are made from a central location, which prevents Marines from manipulating their scores by selecting their supervisors (Cunha et al., 2018).

levels, which increase as they gain credits from institutions of higher education. This is reported as years of education (e.g., an associate's degree is 14 years of education). Increases in formal education are considered in the promotion process, so Marines are incentivized to keep this information updated.

Finally, we track promotions over time. We observe each Marine's job rank on a monthly basis and can trace successive increases in rank over time. We count the number of increases in job rank (i.e., promotions) a Marine achieves relative to 10 months before they have a child. A value of 1 on the variable indicates 1 promotion received since the time point before pregnancy.

Our preferred sample uses a semi-balanced panel of observations to ensure our results are not driven by selective attrition. We require first-time parents (and their potential matches) be observed continuously in the sample for 12 months prior to the birth and 24 months after (n=2,801 mothers). We also require parents to have a military entrance exam score (AFQT or GCT), at least one observed pre-birth fitness score and for parents to have at least one potential nonparent match with the same months of service, rank, and reserve status in the same prepregnancy calendar year. Among mothers, this reduces our sample to n=2,492.

We present characteristics of our first-time parents in Table I, alongside characteristics of first-time civilian parents who are employed and have a child under the age of 1. First-time Marine parents are younger than their civilian counterparts and identify as Black or Hispanic at higher rates than first-time civilian parents, especially mothers. Marine mothers are less likely to be married compared to civilian mothers (68% vs. 81% married), but marriage rates are similar among Marine and civilian fathers (87% and 86% respectively). On average, Marine first-time parents have much lower levels of education than the civilian population of first-time parents. Enlisting as a Marine only a high school degree, so education differences are not unexpected. Fewer than 15% of first-time Marine mothers have attended some college or received a college degree, while more than 87% of civilian mothers who remain employed after having a child have some college or more. The high education rate among civilian mothers may be a function of the type of woman who returns to the labor market after having a child.

We rely on the Standard Occupational Classification (SOC) system from O*NET, a federal standard used to classify workers into occupational categories, to explore the distribution of job types among

Marines relative to civilians. We crosswalk Marine job codes to SOC codes and find that—outside of military-specific occupations—the largest share of first-time Marine fathers work in natural resources, construction, or maintenance (labeled "Construction/maint."), while the largest share of first-time Marine mothers work in sales or office roles. Most civilians who have a first child and stay in the workforce tend to be employed in management, business, science, or arts.

A small share of first-time Marine parents in our sample are officers (akin to civilian managers): 7% and 14% of Marine mothers and fathers, respectively. As such, the vast majority of first births occur to enlisted Marines. Finally, new Marine parents score just above average on required military entrance exams, including the Armed Forces Qualification Test (AFQT) and General Classification Test (GCT).[10] Based on descriptive differences between Marine and civilian first-time parents, results from our analyses may generalize best to younger workers with low levels of formal education.

IV Empirical Approach

The ideal experiment to isolate the causal effect of fertility on men's and women's work performance would randomly assign pregnancy and parenthood to workers. Random assignment would ensure that (on average) differences in outcomes were not driven by underlying characteristics of the types of people who chose to become parents but rather by the transition to parenthood itself. Of course, random assignment of childbirth is both unethical and unfeasible. Yet, a simple post hoc comparison of parents relative to nonparents is unlikely to recover a causal estimate of the effect of having a child. Those who opt into parenthood likely differ from nonparents in ways that might also correlate with work performance.

We use the timing of a first birth to identify the effect of the transition to parenthood. If the transition to parenthood has an impact on work outcomes, then the birth should generate a sharp change in these outcomes at predictable time points. We can attribute any discontinuity in the outcomes at those time points to the pregnancy or birth itself if we assume that other factors that shape job outcomes do not also undergo a sharp change at those same times. In other words, while the choice to have a child may be

[10]Based on the nature of these exams, many officers only have GCT scores and many enlisted only have AFQT scores.

endogenous, the exact timing of conception and subsequent childbirth serves as a shock to the outcomes of interest.

To leverage the shock and precise timing of a birth event, we employ a version of a two-way fixed effects (TWFE) event study strategy. Our goal is to minimize two sources of bias. The first source of bias, which we call TWFE bias, arises when the timing of treatment varies (as it does in our setting) and already-treated units serve as comparison cases for later-treated units, contaminating estimates of counterfactual time trends.[11] To address TWFE bias, Callaway and Sant'Anna (2021) and Sun and Abraham (2021) propose isolating cohorts of units all treated at the same point in time and selecting comparison cases for each treatment time group that *exclude* already-treated units. This strategy functionally aligns event-time and calendar time across treatment and control cases within each treatment time group.

The second source of bias, which we call counterfactual bias, arises if the chosen comparison units do not effectively approximate counterfactual trends in outcomes. In our setting, not all Marines without a birth (or yet to have a birth) would provide a good estimate of counterfactual time trends for those who do have a first birth, especially for promotion outcomes. Promotion eligibility and timing depend on a Marine's current job level and time in service. A second-year enlisted Marine becomes eligible for promotion sooner and can promote faster than a fifth-year officer. Active-duty vs. reserve contexts also give rise to different promotion probabilities. As a result, we face large counterfactual bias if we do not thoughtfully select comparison units for estimates of promotion effects. The same concern around counterfactual bias holds true for estimates of other work outcomes, given that professional expectations for job performance and human capital development depend (though less formally) on job rank, tenure, and active-duty/reserve contexts.

Ideally, our empirical strategy would isolate groups of parents with first births in the same month/year

[11] A traditional TWFE approach incorporates time and unit fixed effects to estimate post-treatment impacts (whether dynamic or constant). Already-treated, not-yet-treated, and never-treated units (if included) contribute to estimates of counterfactual time trends, or the time fixed effects. The estimate of the causal parameter under a traditional TWFE model is then a weighted average of all of two-group/two-period difference-in-difference estimators in the data. As treatment timing gets later, more and more of the two-group/two-period difference-in-difference estimators compare the change in a unit moving from untreated to treated against units who are treated in both the before and after period. This statistical fact requires stronger assumptions than previously recognized, with both a parallel trends assumption *and* an assumption of consistent treatment effects over time and units needed for identification (Goodman-Bacon, 2021). Baker et al. (2021) show that not accounting for the bias in settings that do not meet these assumptions can substantially affect the point estimates of the policy or treatment in question.

(to address TWFE bias) and then draw never-treated comparison units, nonparents, with the same job tenure, job rank, and reserve status as parents (to address counterfactual bias). Under this approach, some groups of parents have few to no nonparent comparison units (especially among reservists and higher ranks). Nonparent comparison units also vary from parents on other important dimensions (e.g., average physical performance, education, age) not defined in the exact match requirements, which generates new concerns around counterfactual bias.

Given this tension, we require that parents and nonparents match on tenure, rank, and reserve status, but we do not focus on exact treatment time groups.[12] Instead, we connect parents to nonparents within the same calendar year. Parents' and nonparents' birth/placebo birth events do not occur in precisely the same month/year, but all observed outcomes are generally happening within the same 12-month timeframe. We hypothesize, in our context, that this method provides the best counterfactual to the treated group. In other words, we are willing to risk some TWFE bias to reduce counterfactual bias. This tradeoff is common in settings where there are a limited number of never-treated or not-yet-treated comparison units once the researcher isolates each cohort of units treated at the same time.

IV.1 Assigning Nonparents to Placebo Births

We use adaptive ridge LASSO (least absolute shrinkage and selection operator) model with 10-fold validation to determine the best predictors of a first birth in our sample. Possible predictors include months of service, an indicator for officer (relative to non-officers), an indicator for reservists (relative to active duty), year, age, race/ethnicity, AFQT scores[13], marital status, an indicator for whether a spouse is also in the military, years of education, occupational field groups, most recent physical performance score, and interactions among all variables. We run this predictive model separately among women and men, measuring all characteristics of first-time parents 10 months before an observed birth (prepregnancy). We then obtain a predicted propensity score for each parent and nonparent.

[12] We also restrict to nonparents we can observe at least 12 months before and 24 months after a potential placebo birth.

[13] We have missing AFQT score data for many officers, who generally take GCT exams. We use a cubic model of standardized GCT scores to predict standardized AFQTs for those with both scores, then use this model to predict AFQT scores for those with only GCT scores.

Among groups of parents and nonparents with exact matches on number of months of service, job rank, reserve status, and observation calendar year, we select up to five nonparents for every parent that are closest in terms of their predicted propensity to have a baby 10 months later.[14] We assign nonparents to a placebo birth event 10 months after the time of the match. Analyses then compare the changes in outcomes for first-time parents to the average change in outcomes for up to 5 most observably similar nonparents to whom they match. Each parent receives a weight of 1 in the analysis, while each nonparent receives a weight of 0.2 per match-month in the case where 5 distinct nonparents match to each parent.

Appendix Table A.1 displays descriptive characteristics of the parents and matched nonparents assigned placebo births, separately by gender. We require matches to be exact on months of service, job rank, reserve status, and calendar year, so the groups exactly match on these first set of characteristics. The next set of characteristics are not exact matches but are included in the LASSO model that predicts the likelihood of first-birth. The groups of first-time parents and nonparents with placebo births look almost identical, with a few small differences within gender. These differences are functionally small (e.g., mothers are 22.57 years old while their placebos are 22.70 years old).

IV.2 Flexible Event Study Estimation

We create a series of event study datasets across exact-match groups (defined by month of service, job rank, reserve status, and calendar year of the pregnancy/placebo pregnancy). We include all month-years of data from before through after the birth/placebo birth, then then stack these exact-match-group datasets.[15] Nonparents assigned placebo births approximate counterfactual time trends in outcomes that first-time parents would have experienced, assuming outcomes would have evolved similarly between parents and those with placebo births. An unbiased estimate in this setting requires an assumption of conditional parallel trends between parents and placebos, which is a weaker assumption than uncondi-

[14] We conduct nearest-neighbor matching with replacement, meaning the same nonparent can be matched to different first-time parents, or parents can match to less than 5 nonparents, who then get higher weights. About 2.8% of mothers and 0.5% of fathers match to fewer than 5 nonparents.

[15] Each month-year observed for a given parent will appear exactly once as they only have one first birth. A given month-year for nonparents may occur multiple times if the same non-parent is matched in different month of service/rank/reserve status/year cell. This matched individual would have different relative event-time points, defined by the time point of their match and assigned placebo birth, for the same observations of calendar month-year.

tional parallel trends had we drawn on all nonparents from the full sample (Roth et al., 2022).

We estimate a fully flexible event study specification separately for men and women as follows:

$$Y_{igtr} = \sum_{r=k_{min}}^{k_{max}} \mathbb{1}(t = t^*_{ig} + r)\theta_r + \pi P_i + \sum_{r=k_{min}}^{k_{max}} \mathbb{1}\left[(t = t^*_{ig} + r) \times P_i\right]\beta_r + \alpha_g + \phi_t + \varepsilon_{igtr} \tag{1}$$

where t^*_{ig} is the month-year of the real or placebo birth for individual i in match group g based on calendar time t. We measure month relative to birth as r. Coefficients θ_r estimate changes in the outcome among match group g for each month r after birth (or before, if $r < 0$) for individual i. This is analogous to event time fixed effects, estimated among both parents and nonparent matches. P_i is a binary variable equal to one for all first-time parents, and we expect π to be null given that parents and placebos are similar in the preperiod. Then, β_r represents how much the parents differ from the their nonparent placebos at a particular time relative to birth. Effects are measured relative to month $r = -10$, which corresponds to 10 months prior to the birth and approximately 1 month before the start of the pregnancy for the parents. We focus on month-by-month effects starting 24 months before birth through 24 months after for most outcomes; for job performance ratings we instead focus on effects starting at $r = -18$ because the evaluation requires workers to be on the job for roughly 6 months before it is typically assessed. We bin event time endpoints below $r = -24$ (or $r = -18$ for job performance) and above $r = 24$. Including binned event time endpoints allows us to estimate time fixed effects ϕ_t to account for month-by-year changes over time in the outcome (e.g., changes in fitness test standards in a particular year) separately from event time fixed effects. We include α_g to create a within-match-group comparison and ε_{igt} as the error term. Thus, θ_{12} represents the average outcome 12 months after the birth for nonparents, relative to $r = -10$, while β_{12} estimates whether this change is larger, smaller, or the same for parents.

Eq. 1 allows us to estimate prepregnancy differences in outcomes between first-time parents and nonparents, reflected by β_r estimates when $r < -10$. Our placebo birth assignment procedure does not require outcomes between parents and nonparents to move together when $r < -10$, but we expect β_r for $r = [-24, -11]$ to be null if nonparents provide a suitable comparison to parents. We show evidence that parents and nonparents are on parallel trends, lending confidence to this estimation strategy.

IV.3 Incorporating Linear Splines

We can improve precision in our estimates of the evolution of outcomes by using a semi-parametric spline specification, especially for estimates with smaller subsamples and outcomes that are not observed every month. Our goal is to identify any level shifts in outcomes during pregnancy, trends during pregnancy, level shifts immediately following birth, and any recovery trends following the immediate impact of birth. Similar to Lafortune et al. (2018) and Bailey et al. (2021), we create a more parsimonious model of changes in outcomes over time by defining:

$Pregnancy_{igtr} = 1$ during months relative to birth $r = [-9, -1]$ and 0 otherwise (for an intercept shift during pregnancy);

$PregnancyTrend_{igtr} = [1, 9]$ corresponding to $r = [-9, -1]$ and 0 otherwise (for monthly trends above and beyond the intercept shift during pregnancy);

$Postbirth_{igtr} = 1$ during months relative to birth $r = [q, 25]$, where $q = 1$ or the earliest time point when the outcome is able to be assessed again after the birth (e.g., starting 8 months after birth for mothers' fitness tests)[16], and 0 otherwise (for an intercept shift following birth);

$Recovery_{igtr} = [q + 1, 25]$ corresponding to $r = [q + 1, 25]$ and 0 otherwise (for monthly trends above and beyond the intercept shift after birth); and

$\Delta Recovery_{igtr} = [1, 13]$ corresponding to $r = [13, 25]$ and 0 otherwise (for any change to the monthly recovery rate that begins at 13 months).

We then modify Eq. 1 to estimate a semi-dynamic specification using these spline parameters:

$$
\begin{aligned}
Y_{igtr} = &\theta_0 Pregnancy_{igtr} + \theta_1 PregnancyTrend_{igtr} + \theta_2 Postbirth_{igtr} + \theta_3 Recovery_{igtr} \\
&+ \theta_4 \Delta Recovery_{igtr} + \pi P_i + \beta_0 (Pregnancy_{igtr} \times P_i) + \beta_1 (PregnancyTrend_{igtr} \times P_i) \\
&+ \beta_2 (Postbirth_{igtr} \times P_i) + \beta_3 (Recovery_{igtr} \times P_i) + \beta_4 (\Delta Recovery_{igtr} \times P_i) \\
&+ \alpha_g + \phi_t + X_{igtr}\gamma_j + (X_{igtr} \times P_i)\delta_j + \varepsilon_{igtr}
\end{aligned} \tag{2}
$$

[16]We define q starting at $q = 1$, the month after the birth, due to ambiguity about whether outcomes measured in the month of birth itself $r = 0$ reflect pre- or postbirth measures.

where effects are measured relative to the prepregnancy average ($r \leqslant -10$), similar to Borusyak et al. (2021).[17] Parents and their matches contribute to all coefficient estimates θ_j, while slope parameters β_j are specific to parents and captures any change above and beyond that of nonparents with placebo births. We present a diagram of this model in Figure II. We use this semi-dynamic spline specification to estimate postbirth effects for men and women at key time points (e.g., 12 and 24 months postbirth). The vector X_{igtr} includes two binary indicators for binned event time endpoints, one for event times $r < -24$ (or $r < -18$ for job performance) and event times $r > 24$ to mirror our estimation strategy from Eq. 1.

Abadie and Spiess (2021) show that clustering should be at the match-group level when doing matching without replacement to account for within-group correlation induced by the matching procedure. That does not solve the problem in matching with replacement that individuals' outcomes are correlated if they are part of multiple matched groups. To address this latter concern, we include two-way clustering at the individual and match-group level.

V Results

V.1 Job Performance

Figure III presents results from our flexible event study model estimated using Eq. 1. We first examine whether outcomes evolve smoothly leading into the pregnancy, suggesting that placebo parents provide a suitable counterfactual estimate of general time trends for the event study sample. The bottom left of each panel includes the p-value of an F-test of whether the prepregnancy point estimates jointly equal zero. We begin with job-related physical performance scores (Panel A). The prebirth period is not jointly statistically different from 0 (p-value of an F-test of joint significance=0.151), indicating that before giving birth mothers' trajectories did not differ from the nonparent women. We exclude outcomes for women during pregnancy and seven months after birth, as policies allowed women to opt out of the physical assessments while pregnant and postbirth. Once women take the test postbirth, their

[17] Using all pretreatment periods is more efficient than using only the period just before the event takes place, but it is also more biased if parallel trends do not hold (de Chaisemartin and D'Haultfœuille, 2021). The event study estimates from Eq. 1 use the period just before treatment as the reference group and thus avoid this issue.

performance declines are large and persistent. Even 24 months postbirth, women's physical performance scores are lower than expected, after accounting for general time trends using the nonparent women. For men, performance declines begin during the mother's pregnancy and reach their lowest point 1 month postbirth. But, the declines are short-lived.

Panel B shows some evidence of lower supervisor ratings of job performance for women in the 2 years postbirth, though estimates are noisier than physical performance given the smaller number of observations. Having a child does not appear to affect fathers' supervisor-rated job performance.

Finally, Panel C indicates that parenthood is unrelated to marksmanship for mothers. Women are exempt from these assessments during pregnancy and while on leave following birth. Given the rarity of this assessment, the estimates are noisy. If anything marksmanship improves postbirth for fathers.

Table II displays the spline results using Eq. 2. This model has the advantage of smoothing the estimates using data in nearby time points to estimate patterns, which is particularly helpful for the noisier event study estimates where outcomes are not observed monthly (i.e., physical performance, job performance and marksmanship). Rather than present coefficients for each spline parameter (e.g., the slope of the postbirth period), we instead predict effects at various time points using coefficients from the model.[18] This can be interpreted as how much parents changed from the prebirth period to the given point in time, net of expected secular trends in the outcome, and whether this change statistically differs from zero.

Beginning at 8 months postbirth, mothers' physical performance is 0.49 standard deviations below their expected average. Mothers recover somewhat, and by 12 months postbirth their predicted physical performance scores are 0.29 standard deviations below expectations. Mothers' physical fitness recovery slows in the child's second year of life. Two years after having a baby, mothers' predicted physical performance remains 0.18 standard deviations lower than before the pregnancy.[19]

Supervisor ratings do not change during pregnancy (see Appendix Table A.3). One month postbirth,

[18] We include the slope parameters in Appendix Table A.3 for reference.

[19] Physical performance assesses a combination of cardiovascular health, endurance, and upper-body and core strength. Appendix Table A.4 presents raw scores for each item on the fitness assessments that measure overall physical performance. Declines in women's performance occur across almost all measured assessments. The initial postbirth drop shows women run more slowly (i.e., run times increase) and complete fewer crunches and pull-ups/lifts.

mothers score 0.17 standard deviations lower than expected, relative to changes in the placebos. This gap is 0.18 standard deviations at one year postbirth and a non-statistically significant 0.07 standard deviations by 24 months postbirth. Marksmanship is lower for mothers than their placebos postbirth, though by 24 months postbirth the difference is no longer statistically significant.

The patterns for fathers' job-relevant physical performance is consistent with mothers' but smaller in magnitude (Panel B of Table II). We see an immediate drop in performance of 0.12 standard deviations in the month after birth. By 12 months postbirth, the fathers are 0.05 standard deviations below expectations, and the effect is a precise zero by 24 months postbirth. For job performance, scores are slightly negative in the first month postbirth, but return to match the placebos by the child's first birthday. For marksmanship, if anything fathers may improve over time.

V.2 Training, Education, and Promotion

We next turn to human capital development and promotion outcomes. These outcomes are recorded monthly, meaning every sample member has an observation for every month-year before and after the birth, giving us more precision in our estimates. Figure IV presents results from Eq. 1; Figure A.6 displays the unadjusted weighted means over time for parents and their placebos.

Figure IV Panel A shows impacts on months of job-specific training, where the Y-axis measures the total count of months in training relative to $r = -10$. The prepregnancy point estimates do not jointly different from zero, but there appears to be a downward trend, particularly among mothers. Yet, starting in pregnancy mothers' accumulation of months spent in training slows even more relative to nonmothers'. The gap does not close after the birth. Table II shows that the gap grows from 0.46 months immediately after birth to 0.84 months 24 months postbirth, which is about an 51.9% difference from the nonparent mean growth of 1.61 months in $r = [-10, 24]$. There does not appear to be a meaningful impact of birth on accumulated job-specific training among fathers.[20]

[20]To address concerns around pretrends in the training outcome, we control for a linear pretrend and show predicted impacts in Appendix Table A.2. The direction of the effects for months of training among mothers is similar, but the magnitude of the impacts is slightly smaller. Among fathers, controlling for pretrends in months of training produces results that suggest, if anything, fathers' time spent in training increases relative to nonfathers after the birth. We interpret this result with caution, given that the estimates are sensitive to the specification we use.

Figure IV Panel B shows the impact of a first birth on total years of formal education, a transferable measure of human capital. Mothers have slightly lower educational attainment than their placebos by 24 months postbirth, but there is some evidence of pretrends in Figure IV; when we control for pretrends in Appendix Table A.2 we do observe any impact of a birth on mothers' total years of education. Among fathers, there is a statistically significant but practically small increase relative to the placebos of 0.01 years of education 24 months following the birth. However, we do not observe this effect when we control for pretrends in the outcome. Our overall conclusion is that birth has minimal effects on working parents' educational attainment in our context.

The final row in Figure IV displays promotions, which is a cumulative count relative to $r = -10$. Mothers and their placebos move together before pregnancy, but a gap emerges around the time of the birth. Mothers never catch up to their placebos; if anything the gap grows over time. From Table II, the promotion gap is 0.03 at $r = 1$ but grows to 0.09 promotions by $r = 24$. The placebos averaged 1.31 promotions in $r = [-10, 24]$, so this is an 6.7% reduction. Fathers are statistically about the same as their placebos on this outcome.

V.3 Heterogeneity across Subgroups

Parenthood may affect groups differently in ways that may either uncover what we think may happen in others (non-military) contexts or in ways that may be of interest for their own sake. We explore heterogeneity in Figure V; job performance and marksmanship scores are in Appendix Figure A.7 given their lower power. We conduct this analysis by interacting an indicator variable for a characteristic (e.g., reservist) with the variables from Eq. 2; this measures whether the change in the gap between parents and nonparents from prepregnancy to postpregnancy differs across group types.[21] In general, we lack the power to test out small differences between subgroups of mothers, but are well-powered to detect even small differences among subgroups of fathers. Therefore, we focus mainly on statistically significant

[21] Definition of the subgroup is based on the matched parents' characteristics for the placebos, rather than the characteristics of the placebos themselves. This ensures the placebos are in the same group as their matched parents, even if they don't exactly match on a characteristic. We use $r = -10$ to define most subgroups. We define marital status at $r = 0$ because we are most interested in family characteristics when the baby arrives. This time gives couples surprised by a pregnancy some time to marry if they want to.

differences among groups of fathers at the 1% level. Cross-group differences that differ at the 1% level are shown by the filled-in markers in Figure V.

We first consider variation in impacts between reserve and active-duty Marines. Reservists provide a useful comparison to active-duty service members, as most reservists have full-time civilian jobs but have the same physical, job performance, and marksmanship requirements as active-duty. If we see larger (or smaller) impacts of birth among reservists, it may give a sense of how civilians differ from those who make their primary living from being in the military.[22] For fathers, the drop in physical performance immediately following birth is much larger for reservists than active-duty service members (p-value of difference=0.002). However, by 24 months postbirth, the size of the impact does not differ between active-duty and reserve fathers for any outcome. We may then expect that physical performance would drop even more in civilian settings in response to parenthood, if individuals' primary jobs are not explicitly evaluated on physical performance.

A potential concern in our analysis is that Marines who plan to leave the military soon after we observe them postpartum may not be as invested in their performance. To address this, we next split the sample by those who stay in the Marines for 36 months or longer postbirth (75% of parents) and those who leave in months 25–35. For fathers, those who leave earlier consistently experience more negative impacts of parenthood than those who stay in; these differences are statistically significantly different at p=0.000 for physical performance, training, education, and promotion and by $r = 24$. Moreover, the mothers' promotion gap at $r = 24$ is driven by those about to leave (p=0.000).

Larger negative effects for parents who leave sooner could be because they wanted to get out, put in less effort at work, and then did not get promoted – or because they struggled to perform after having a child, did not get promoted, and decided to leave as a result. We cannot disentangle these two possibilities. Still, even among those who stay in, we generally find the same patterns as the main analysis: worse physical performance, less training, and (slightly) lower number of promotions for mothers, and a short-term drop in physical performance but otherwise limited (or potentially even positive) effects for

[22] We do not have the power to conduct the reservist subgroup analyses for physical performance, job performance, and marksmanship scores for women due to only a small number of female reservists and sporadic nature of the assessments. Reductions in training after birth are concentrated among active-duty mothers (p=0.000 at both time points)

fathers.

Parents in high-physicality jobs may have different incentives to stay fit—or more ground to lose in terms of their physical ability on the job. We categorize Marines as working in a "high-physicality" ($> physical$) or "low-physicality" ($< physical$) job if their job responsibilities place them in the top half or bottom half of the distribution of O*NET's physicality index. We exclude individuals working in jobs that have no link to an O*NET classification (about 9% of mothers and 32% of fathers). Fathers in higher-physicality jobs have relatively larger drops in physical performance than those in lower-physicality jobs at $r = 1$ (p=0.000), while those in lower-physicality jobs have relatively low training completion at $r = 1$ and $r = 24$ (p=0.002 and 0.007, respectively). For education and promotion, fathers in lower-physicality jobs outperform their placebos, while high-physicality fathers match their placebos; these patterns are statistically significantly different at $r = 1$ and $r = 24$ for education and promotion (p=0.000 for all comparisons). Overall, we have evidence that it is more difficult for fathers in physically demanding jobs to maintain their performance. The promotion gap is driven by mothers in higher-physicality jobs (p=0.005 for $r = 1$ and p=0.000 for $r = 24$).

Unmarried parents may have less support at home and may struggle more at work. Single fathers display lower performance on the fitness test than married fathers at $r = 1$ (p=0.010). The gap is worse for single fathers at 24 months for training (p=0.000) and education (p=0.007), but not promotions. Single mothers have larger promotion effects than married mothers at $r = 24$ (p=0.002). Overall, we have evidence that the negative impacts of childbirth on performance and human capital outcomes are larger for single parents than for their married counterparts. Junior enlisted may be less attached to their jobs than senior enlisted or officers, or they may be less able to advocate for themselves as lower-ranking members of the organization. For fathers, the negative impact of parenthood on physical performance is larger for junior enlisted fathers than for more senior-ranking fathers at $r = 1$ (p=0.000). More senior fathers actually outperform their placebos and receive more training and education in response to childbirth, while junior enlisted fathers receive less (p=0.000 for both training and education). To some extent, then, fathers "rise to the challenge" of fatherhood by outperforming otherwise comparable non-fathers—but only the more senior-ranking fathers.

For mothers, junior enlisted drive the long-term lower physical performance at $r = 24$ (p=0.008). Conversely, senior women have larger declines in time spent in training at $r = 1$ and $r = 24$ (p=0.001 and 0.004, respectively). It is junior enlisted who drive declines in promotions at $r = 1$ and $r = 24$ (p=0.008 and 0.000, respectively); promotion paths of senior mothers do not differ from their placebos.

V.4 Mechanisms Behind Delays in Mothers' Promotion Trajectories

Given that mothers have promotion gaps relative to comparable nonmothers, an immediate question is whether changes in job performance and human capital development drive this difference in promotions. We consider four primary pathways by which motherhood could translate into promotion gaps. First, drops in performance on the job could explain slower promotion. Second, changes in human capital development could translate into delayed promotion, especially if mothers need certain training to reach the next promotion. Third, pregnant and postpartum mothers have the option of waiving physical fitness testing; this is why we excluded $r = [-9, 7]$ in the main analysis. Promotion boards may view missed assessments negatively; female Marines we talked to believed missing fitness tests could be detrimental at promotion time. Indeed, some mothers still opt into taking the fitness test early upon return to work and others conceal their pregnancies in order to continue testing. Variation also exists for nonmothers in job performance, human capital development and number of completed fitness tests (for nonmaternity-related reasons). For instance, some nonmothers get fitness test waivers for medical reasons such as spraining an ankle. Finally, a fourth pathway is that there is simply discrimination against mothers, and none of the observed characteristics fully explain the promotion gaps.[23]

We do a simplified mediation test of these pathways in Table III. The outcome is the count of promotions accumulated in $r = [-10, 24]$. We include mothers, fathers, and their respective placebos in the same regression with each person-match entered once. All models control for baseline characteristics and include an indicator for being female, an indicator being a parent, and an interaction between these two.[24] The coefficient on this interaction would be negative if mothers have low promotion levels, above

[23] We note that even if we do fully explain mothers' promotion delays, discrimination may still exist. Supervisor's job performance ratings may be biased, and assessments may systematically disadvantage women if they, say, test abdominal strength soon after childbirth, even if every Marine has to do crunches.

[24] Parents have a weight of 1; placebos have the weight from the main analysis. Control values are taken at $r = -10$

and beyond being a parent in general, being a female in general, and the other variables included in the model. Column (1) provides a baseline. Fathers in general have 0.02 more promotions than predicted by their baseline characteristics; mothers' promotions are much lower than comparable females.

Next, we assess whether various interim outcomes explain this gap. Column (2) adds a variable for the mean fitness scores for $r > 0$. The coefficient on the mother interaction shrinks by 28% (see Sobel-Goodman mediation), implying that changes in physical performance explain roughly a quarter of the motherhood promotion gap.[25] Job performance, marksmanship, training, and education (Columns 3–6) do not explain the promotion gap. Entering all of these variables simultaneously explains about a quarter of the motherhood promotion gap (Column 7).

Physical fitness performance matters for promotion, but mothers systematically miss these assessments. We next examine whether missing physical fitness assessments changes promotion rates. We operationalize this measure by counting how many of the 6-month fitness cycles that ended in $r > -10$ each individual missed. We enter this as a total count per person-match, which averaged 2.6 waived tests for mothers, and had a range of waived tests for placebo women, fathers, and placebo fathers. Having more waived fitness tests is associated with lower rates of promotion for everyone, and this variable explains all of the motherhood gap (Column 8); patterns are similar when all potential mechanisms are included (Column 9). Overall, missing evaluations seems to be particularly important for explaining why mothers are not promoted.[26]

and include months of service, an indicator for officer, an interaction between months of service and officer, and the other variables from the matching model.

[25]The Sobel-Goodman mediation test examines whether a mediator (e.g., fitness scores) carries the influence of some explanatory variable (e.g., motherhood) to the dependent variable (e.g., promotion). The test specifically computes the proportion of the total mother-specific effect of parenthood on promotion counts that is mediated by the additional sets of controls.

[26]Appendix Table A.5 explores several additional pathways: being deployed, months of being in a nondeployable status, and the average of whether a Marine achieved a "first-class" fitness score rather than the continuous measure to measure fitness. Actually deploying does not have a large mediation effect. Being in a limited duty status (which is strongly associated with waiving fitness tests) for medical or other reasons accounts for over half of the motherhood promotion penalty. Measuring fitness as attaining a first-class score does a slightly better job of mediating the motherhood penalty than do the standardized scores, possibly because promotion boards examine fitness in this dichotomized way.

V.5 Variation by Maternity Leave Length

Prior to 2015, all DoD branches provided active-duty women with 6 weeks of paid maternity leave. In July 2015, the Secretary of the Navy announced that primary caregivers in the Navy and Marine Corps would be entitled to 18 weeks of leave. Women who had given birth earlier in the year (as of January 2015) could retroactively take advantage of the 18-week leave policy before their child turned one. Women who had already returned to work after 6 weeks of leave tended to use the additional 12 weeks of paid time off discontinuously (i.e., as flexible time off). Women who were on leave at the time of the announcement of expanded leave, or gave birth after the announcement, generally took the additional leave consecutively (Bacolod et al., 2021). We analyze these groups separately, referring to the different leave arrangements as "6 weeks + 12 flex" to indicate discontinuous extended leave used as flexible time off, and "18 weeks" to signal the stretch of continuous extended weeks of leave. In early 2016, the Secretary of Defense standardized maternity leave to 12 weeks for all military branches. The 12-week policy applied to pregnancies that began 31 days after the announcement (i.e., pregnancies that began on March 3, 2016 or later, per doctor estimation).

We disaggregate the effects of having a child on women's outcomes by the length of maternity leave, determined by when she gave birth. The key question of interest is whether longer or shorter leave predicts better or worse outcomes when women return to work. Variation in leave length is, at times, quasi-randomly assigned. Some policy changes were unexpected and applied to women who were already pregnant, while other changes were prospective, allowing women to potentially select into parenthood and a particular leave policy. Selection presents the biggest concern for women who gave birth under the latter part of the 18-week policy and the full 12-week policy, given that these women would have known their leave length before becoming pregnant. Instead, we focus on women who expected to receive 6 weeks of leave at conception and were surprised with the announcement of additional leave. We compare three distinct lengths of leave: 6 weeks, 6 weeks + 12 flexible weeks after returning to work, and 18 weeks. Appendix Table A.6 presents descriptive characteristics of the women who gave birth under these three leave policy groups. There are some differences across groups, though not in any ways that suggest systematic bias in one policy regime or another.[27] We conduct the analysis by defining indicator variables for the "6 weeks + 12 flex" and "18 weeks," with the "6 week" policy as the baseline group.[28] We interact these indicators with the variables in Eq. 2 and make policy-specific predictions for the initial birth drop (i.e., 8 months for physical performance and 1 month for training, education, and promotion), 12 months postbirth, and 24 months postbirth.[29]

For reference, Table IV replicates the main analysis for the subsample who expected 6 weeks of leave at conception (and their matches). Each panel then shows the results from the regression with policy interactions. Among mothers in our three policy groups of interest, physical performance drops when initially observed 8 months after having a baby (as it does in the full sample). Mothers who had longer maternity leave had larger physical performance declines, particularly those under the "6 weeks + 12 flex" policy. These mothers had returned to work following their initial 6 weeks of leave, then received an additional 12 weeks of leave they had to use by their child's first birthday. However, an F-test indicates that the magnitude of the declines across leave policies do not statistically differ from each other at the 5% level (p(diff), all effects=0.056). Results from all other F-tests do not indicate that the impact of parenthood on work outcomes meaningfully varies based on leave length. The only statistically significant difference is for education, where mothers with more leave experience larger declines in years of formal education by $r = 24$ after having a child.

[27] There are significant differences in percent who are officers, years of education, and combat job type. We also test whether the density of first births is continuous across each policy threshold, providing additional evidence on possible selection. Figure A.8 shows month-by-month variation in the density of births, including a test for any discontinuity across policy thresholds, following Cattaneo et al. (2018). None of the differences across the policy thresholds reach statistical significance, suggesting the policies did not influence female fertility itself.

[28] In supplemental analyses we set aside our concern about selection into having a birth and include mothers who knew they would receive additional leave at conception; that analysis includes more observations in the "18 weeks" period and an additional group of women under the "12 weeks" of leave period. The balance table for this sample is in Appendix Table A.7, and the results are in Appendix Table A.8. Because babies born in November–December 2016 could have fallen under either the 18 or 12 week policy depending on date of conception, we exclude these mothers (and their matches) from the policy analysis.

[29] We do not include supervisor-rated job performance or marksmanship here. Supervisor ratings for young enlisted Marines are only available for those who remained in service as of October 2017, which complicates analyses of policy changes that took place in 2015 and 2016. In terms of marksmanship scores, we lack power to subdivide this outcomes given sparse observations.

V.6 Alternative Specifications

Our preferred empirical strategy prioritizes identifying an appropriate set of comparison cases to model counterfactual trends that first-time parents would have experienced absent a birth. Tables V and VI explore several alternative models for mothers and fathers, respectively. The table columns are our main outcomes of interest, while each row in a given segment shows results for alternative specifications.

Row (1) is a traditional TWFE event study. The comparison group is all same-gender Marines who do not have a baby and remain in the Marine Corps at least 3 years. We are not worried about TWFE bias because we have many control individuals, meaning that most of our overall estimates will come from comparisons of untreated-to-treated against always-untreated controls. The concern with this model is that the parents are not the same as the average nonparent and thus nonparents do not provide a helpful counterfactual. We cannot use the TWFE approach to estimate training or promotion impacts because we measure these outcomes cumulatively, and the cumulative measure requires a pregnancy starting point that the comparison group (not assigned a placebo birth) does not have.

Row (2) is a stacked TWFE model, which identifies cohorts of units treated at the same time, excludes any already-treated units from each cohort, stacks the cohorts, and then runs TWFE models across cohort-specific groups.[30] Each parent is connected to five nearest neighbor nonparents from the LASSO prediction model such that parents' births and nonparents' placebo births occur in the same month and year, which eliminates the negative weighting that can occur in traditional TWFE models (Callaway and Sant'Anna, 2021; Cengiz et al., 2019; Sun and Abraham, 2021). The LASSO model used to match parents and nonparents is the same as in the preferred analysis. However, we do not require an exact match between parents and nonparents on months of service, rank, or active/reserve status, which may generate counterfactual bias (i.e., where nonparents do not approximate counterfactual outcomes well).

Row (3) is similar to our preferred model, but it only includes data from $r = [-24, 24]$ and removes the binned endpoints at $r = -25$ and $r = 25$. The binned endpoints were necessary for modeling both time relative to the match and month-year fixed effects. The model in Row (3) removes the month-year

[30]In a stacked approach, regression estimates from each treatment-time cohort are combined using variance weighting to recover a single estimate of the impact across cohorts. Recently proposed alternative estimators, for example by Callaway and Sant'Anna (2021) and Sun and Abraham (2021), use other approaches to weighting each cohort-specific treatment estimate.

fixed effects, which requires the assumption that the relative time trends in Eq. 2 sufficiently capture counterfactual time trends. We use the same sample of individuals as in the preferred specification.

Row (4) is our preferred model provided for reference. It includes binned endpoints, month-year fixed effects, and exact matching on months of service, rank, active/reserve status, and calendar year.

The top panels of Tables V and VI show a series of F-tests assessing the pretrends in $r = [-24, -10]$ by outcome for the different models. This is analagous to the F-test displayed in Figures III and IV. If the untreated group is a good counterfactual, we would expect these estimates to be null. There is some evidence of pretrends for both placebo estimates in education, with a p-value of 0.030 for the model with no time fixed effects and 0.025 for the preferred model with time fixed effects. Given the number of outcomes we examine, this could happen by chance, but for this reason we take the women's education outcomes with a grain of salt. All of the other exact match pretrends are null for the mothers. The standard TWFE models is more precisely significant for the pretrends in education ($p = 0.000$), while the stacked TWFE model is significant for physical performance ($p = 0.009$), marskmanship ($p = 0.023$), and training ($p = 0.000$). The fathers show some evidence of pretrends in 25% of the standard TWFE outcomes, 67% of the stacked TWFE outcomes, and none of the two matched placebo outcomes. This highlights the importance of the exact matching in our setting.

The middle and bottom panels of the tables show the predicted value for the given outcome at $r = 12$ and $r = 24$, respectively. Both exact-match placebo birth strategies (with and without time fixed effects) produce almost identical results, and the broad takeaways are generally consistent across all models for women: there are large drops in physical performance that never return to pre-pregnancy levels, while training, education, and promotions remain below expectations at $r = 24$. However, the size of these predictions differ by model and highlight the importance of choosing the best comparison group. The parallel trends assumption means that the nonparents represent a good counterfactual to the parents in the postperiod. Parallel pretrends offers support for this assumption, but parallel pretrends do not guarantee parallel postreeatment counterfactual trends. In our case, parents' (unobserved) counterfactual postbirth trajectory may differ from nonparents' (observed) postbirth trajectory. As an example, women in combat roles may have better expected physical fitness trajectories in the long-run but are also less likely to become mothers. These women in combat jobs would not be a good counterfactual to the average mothers. For this reason, we prefer the conditional parallel trends assumption required in the stacked and exact matching strategies.

We know that promotion is mechanically tied to rank and time in service, so it is particularly important to ensure parents and nonparents match on these characteristics in the preperiod. We would not want to match a low-ranking officer to mid-ranking enlisted, even if they have similar rates of promotion in the preperiod, as their subsequent expected promotion trajectories differ even in the absence of a child. Indeed, when comparing the stacked fixed effect model to the preferred exact-match model, the promotion gap is 55% larger for mothers and, for fathers, flips direction and becomes statistically significant.

Our approach that priotizes defining an appropriate counterfactual while still considering how to minimize TWFE bias may be useful in cases where cells sizes are too small to isolate distinct cohorts of cases treated at the same time and maintain a sufficient number of comparison cases, especially in settings where other variables (e.g., months of service) are particularly important to consider for creating a counterfactual group. For instance, with yearly state-level data where total observations are limited by the total number of U.S. states, researchers might consider grouping proximate treatment years rather than defining groups of states with the same precise treatment year.

VI Summary and Conclusions

We use repeated, direct measures of work performance, human capital accumulation, and career advancement to explore the link between the transition to parenthood and workers' outcomes. Our empirical strategy draws on an event study approach based on the timing of a first birth and a matching design that assigns placebo births to observably similar nonparents. We find both men and women's health-related job performance responds negatively to the transition to parenthood. However, gender differences emerge in the magnitude and persistence of decline and recovery postbirth. Women experience large declines in job-related performance that can persist for two years postbirth, while men experience smaller, short-lived declines in job performance that fade by their child's second birthday. Mothers' accumulated time spent in on-the-job training slows, while fathers' outcomes are largely unaffected, if not slightly improved. These patterns are consistent with our findings that women's promotion trajectories slow after having a child while men's do not. Among women, promotion delays accumulate over time; the gap in number of promotions between mothers and nonmothers is largest 24 months postbirth.

Gender-differences in the promotion effects of parenthood directly lead to pay gaps in this setting. By 24 months postbirth, the average mother would make $40,596 in basic pay according the Marine Corps pay schedule (excluding any bonuses or housing allowances).[31] The impact of birth on promotions means that mothers go from an average of $0 difference in pay compared to matched nonmothers 10 months before having a first child to $332 lower pay at 24 months postbirth. The female-male wage differential grows from $5,789 10 months before women or men have a first child to $5,890 at 24 months postbirth (a $101 increase). Fathers are generally more advanced in their careers before they have a child, which generates the prepregnancy gender-based wage differential. This difference in career advancement highlights the importance of using similar "placebo" parents for the comparison: we would systematically underestimate maternal wage penalties using fathers as a comparison because mothers are earlier in their careers (when pay is lower and promotions *should* happen faster) than fathers. This finding also contributes more broadly to the idea that family fixed effects models may not provide a useful design if other members of a family are also affected by a given event (Black et al., 2021; Cools and Patacchini, 2019; Fletcher and Marksteiner, 2017; Heissel, 2021).

We explore potential mechanisms behind observed gender differences in accumulated promotions after having a child. We find some evidence that delays in promotion are driven by mothers' postpartum exemptions from key evaluations used in determining promotions, as well as lower physical performance.

Last, we show that additional leave does not improve mothers' work-related outcomes, but it largely does not exacerbate the negative impact of parenthood, either. If the goal of maternity leave is to provide bonding time with children and time for mothers to physically, medically, and mentally recover, this may be good news: job-related motherhood penalties were not exacerbated in this context by more generous

[31] We use 2022 basic pay scales for this estimate. Basic pay is calculated by years of service and rank. Marines also have housings allowances that increase with number of dependents (thus counteracting the mothers' gap) and also increase with rank (thus increasing the gap if promotions are delayed), as well as other bonuses or incentives that may differ for mothers (e.g., combat pay). We focus on basic pay because it aligns most closely to civilian pay and is straightforward to calculate.

parental leave policies, especially in the longer-term, two years after the birth.

Our findings provide a new angle on the longstanding literature that shows parenthood reduces mothers' employment, hours worked, and wages, while having no effect on fathers. We find having a child impacts mothers' job performance and skill accumulation in the first two years of the child's life, highlighting the period after birth as a possible critical window that could give rise to long-term child penalties. Delays in promotion that accumulate for women, but not for men, in the years following birth also underscore the need for increased policy- and firm-level support for recent parents. Future research could explore whether family support policies, such as increased access to affordable child care, mitigate challenges parents face as they transition back into work after having a child.

Author Affiliations

Olivia Healy is a postdoctoral fellow in the Department of Policy Analysis and Management at Cornell University in Ithaca, NY (ojh4@cornell.edu). Jennifer Heissel is an assistant professor in the Department of Defense Management at the Naval Postgraduate School in Monterey, CA (jaheisse@nps.edu).

References

Abadie, A. and Spiess, J. (2021). Robust post-matching inference. *Journal of the American Statistical Association*, pages 1–13.

Aguilar-Gomez, S., Arceo-Gomez, E., and De la Cruz Toledo, E. (2019). Inside the black box of child penalties. *Available at SSRN 3497089*.

Agüero, J. M. and Marks, M. S. (2011). Motherhood and Female Labor Supply in the Developing World: Evidence from Infertility Shocks. *The Journal of Human Resources*, 46(4):800–826.

Andresen, M. E. and Nix, E. (2020). What Causes the Child Penalty? Evidence from Same Sex Couples and Policy Reforms. page 69.

Angelov, N., Johansson, P., and Lindahl, E. (2016). Parenthood and the Gender Gap in Pay. *Journal of Labor Economics*, 34(3):545–579.

Angrist, J. D. and Evans, W. N. (1998). Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size. *The American Economic Review*, 88(3):450–477.

Azmat, G. and Ferrer, R. (2017). Gender gaps in performance: Evidence from young lawyers. *Journal of Political Economy*, 125(5):1306–1355.

Bacolod, M., Heissel, J. A., Laurita, L., Molloy, M., and Sullivan, R. (2021). Mothers in the military: The effect of maternity policy on leave up-take. *Demography*. In press.

Bailey, M., Byker, T., Patel, E., and Ramnath, S. (2019). The Long-Term Effects of California's 2004 Paid Family Leave Act on Women's Careers: Evidence from U.S. Tax Data. Technical Report w26416, National Bureau of Economic Research, Cambridge, MA.

Bailey, M. J., Sun, S., and Timpe, B. (2021). Prep school for poor kids: The long-run impacts of head start on human capital and economic self-sufficiency. *American Economic Review*, 111(12):3963–4001.

Baker, A., Larcker, D. F., and Wang, C. C. (2021). How much should we trust staggered difference-in-differences estimates? *Available at SSRN 3794018*.

Balser, C., Hall, C., and Bukowinski, A. (2020). The Effect of Expanding Paid Maternity Leave on Maternal Health: Evidence From the United States Air Force & Army. SSRN Scholarly Paper ID 3601018, Social Science Research Network, Rochester, NY.

Barth, E., Kerr, S. P., and Olivetti, C. (2017). The Dynamics of Gender Earnings Differentials: Evidence from Establishment Data. Technical Report w23381, National Bureau of Economic Research, Cambridge, MA.

Bertrand, M., Goldin, C., and Katz, L. F. (2010). Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors. *American Economic Journal: Applied Economics*, 2(3):228–255.

Black, S. E., Breining, S., Figlio, D. N., Guryan, J., Karbownik, K., Nielsen, H. S., Roth, J., and Simonsen, M. (2021). Sibling spillovers. *The Economic Journal*, 131(633):101–128.

Borusyak, K., Jaravel, X., and Spiess, J. (2021). Revisiting event study designs: Robust and efficient estimation. *arXiv preprint arXiv:2108.12419*.

Bronars, S. G. and Grogger, J. (1994). The Economic Consequences of Unwed Motherhood: Using Twin Births as a Natural Experiment. *The American Economic Review*, 84(5):1141–1156.

Bullinger, L. R. (2019). The effect of paid family leave on infant and parental health in the united states. *Journal of health economics*, 66:101–116.

Bureau of Labor Statistics (2017). Physical strength required for jobs in different occupations in 2016. Bureau of Labor Statistics and United States Department of Labor.

Butikofer, A., Riise, J., and Skira, M. (2018). The impact of paid maternity leave on maternal health. *NHH Dept. of Economics Discussion Paper*, (04).

Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230.

Cattaneo, M. D., Jansson, M., and Ma, X. (2018). Manipulation testing based on density discontinuity. *The Stata Journal*, 18(1):234–261.

Cengiz, D., Dube, A., Lindner, A., and Zipperer, B. (2019). The effect of minimum wages on low-wage jobs. *The Quarterly Journal of Economics*, 134(3):1405–1454.

Cools, A. and Patacchini, E. (2019). The brother earnings penalty. *Labour Economics*, 58:37–51.

Cools, S., Markussen, S., and Strøm, M. (2017). Children and Careers: How Family Size Affects Parents' Labor Market Outcomes in the Long Run. *Demography; Silver Spring*, 54(5):1773–1793.

Cortes, P. and Pan, J. (2020). Children and the remaining gender gaps in the labor market. Technical report, National Bureau of Economic Research.

Cruces, G. and Galiani, S. (2007). Fertility and female labor supply in Latin America: New causal evidence. *Labour Economics*, 14(3):565–573.

Cunha, J. M., Shen, Y.-C., and Burke, Z. R. (2018). Contrasting the impacts of combat and humanitarian assistance/disaster relief missions on the mental health of military service members. *Defence and Peace Economics*, 29(1):62–77.

Cáceres-Delpiano, J. (2006). The Impacts of Family Size on Investment in Child Quality. *The Journal of Human Resources*, 41(4):738–754.

de Chaisemartin, C. and D'Haultfœuille, X. (2021). Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey. *Available at SSRN*.

Department of Defense (2018). 2018 demographics report: Profile of the military community". Technical report, Department of Defense (DoD) Office of the Deputy Assistant Secretary of Defense for Military Community and Family Policy under contract with ICF.

Fletcher, J. and Marksteiner, R. (2017). Causal spousal health spillover effects and implications for program evaluation. *American Economic Journal: Economic Policy*, 9(4):144–66.

Gallen, Y. (2018). Motherhood and the gender productivity gap. *Becker Friedman Institute for Research in Economics Working Paper*, (2018-41).

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*.

Heissel, J. A. (2021). Teen fertility and siblings' outcomes evidence of family spillovers using matched samples. *Journal of Human Resources*, 56(1):40–72.

Jacobsen, J. P., Pearce, J. W., and Rosenbloom, J. L. (1999). The Effects of Childbearing on Married Women's Labor Supply and Earnings: Using Twin Births as a Natural Experiment. *The Journal of Human Resources*, 34(3):449–474.

Kim, S. and Moser, P. (2020). Women in science: Lessons from the baby boom. National Bureau of Economic Research Summer Institute (NBER SI) Session on Gender in the Economy.

Kleven, H., Landais, C., Posch, J., Steinhauer, A., and Zweimüller, J. (2020). Do family policies reduce gender inequality? evidence from 60 years of policy experimentation. Technical report, National Bureau of Economic Research.

Kleven, H., Landais, C., Posch, J., Steinhauer, A., and Zweimüller, J. (2019a). Child Penalties Across Countries: Evidence and Explanations. NBER Working Paper w25524, National Bureau of Economic Research, Cambridge, MA.

Kleven, H., Landais, C., and Søgaard, J. E. (2019b). Children and Gender Inequality: Evidence from Denmark. *American Economic Journal: Applied Economics*, 11(4):181–209.

Lafortune, J., Rothstein, J., and Schanzenbach, D. W. (2018). School finance reform and the distribution of student achievement. *American Economic Journal: Applied Economics*, 10(2):1–26.

Pac, J. E., Bartel, A. P., Ruhm, C. J., and Waldfogel, J. (2019). Paid family leave and breastfeeding: Evidence from california. Technical report, National Bureau of Economic Research. NBER Working Paper.

Roth, J., Sant'Anna, P. H. C., Bilinski, A., and Poe, J. (2022). What's trending in difference-in-differences? a synthesis of the recent econometrics literature. Papers, arXiv.org.

Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199.

Trajkovski, S. et al. (2019). California paid family leave and parental time use. Technical report, Center for Policy Research, Maxwell School, Syracuse University.

VII Tables and Figures

Table I: Characteristics of First-Time Parents

	Mothers		Fathers	
	Marines	Civlian	Marines	Civilian
Descriptive characteristics				
Age	23.40	29.97	25.47	31.80
Black	0.15	0.09	0.10	0.07
Hispanic	0.23	0.12	0.14	0.14
Married	0.68	0.81	0.87	0.86
Some college	0.05	0.28	0.05	0.27
College	0.09	0.59	0.16	0.48
Job Classifications				
Mngmt./Business/Science/Arts	0.13	0.57	0.10	0.46
Service	0.07	0.15	0.04	0.11
Sales/Office	0.35	0.24	0.12	0.15
Construction/Maint.	0.18	0.00	0.29	0.15
Production/Moving/Transpo.	0.19	0.03	0.14	0.14
Military	0.08	0.00	0.31	0.00
Military-Specific Characteristics				
Officer	0.07	–	0.14	–
AFQT score (percentile)	58.56	–	63.23	–
GCT score (av=100; sd=20)	103.38	–	111.29	–
N of individuals	2,492	3,638,695	24,066	4,557,719

Notes: Displays characteristics of first-time parents in the Marine Corps in our sample alongside characteristics of first-time civilian parents in the labor market. Time-varying characteristics of Marines in our sample (e.g., age) are measured 10 months before the birth (t=-10). Data on civilians come from the American Community Survey 1-year estimates, 2010 to 2018. We limit the civilian sample to adults who are employed in the civilian labor market and have a first child under age 1. Job categories correspond to Standard Occupational Classification (SOC) system groups applied to U.S. Marine Corps job codes and available in the American Community Survey. Military specific variables include whether a Marine is ranked as an officer (akin to manager) and AFQT and GCT scores, which are measures of intelligence. We do not observe these military-specific variables in the civilian sample.

Table II: Impacts of Childbirth Among First-Time Parents

	Physical Performance (sd)	Job Performance (sd)	Marksmanship (sd)	Training (months)	Education (years)	Promotions (count)
	(1)	(2)	(3)	(4)	(5)	(6)
			A. Mothers			
1-month effect	–	-0.174***	–	-0.410***	-0.015	-0.032**
		[0.000]		[0.000]	[0.086]	[0.003]
8-month effect	-0.494***	-0.178***	-0.118**	-0.602***	-0.022*	-0.058***
	[0.000]	[0.000]	[0.005]	[0.000]	[0.017]	[0.000]
12-month effect	-0.290***	-0.181***	-0.132**	-0.712***	-0.027*	-0.074***
	[0.000]	[0.000]	[0.006]	[0.000]	[0.012]	[0.000]
24-month effect	-0.183***	-0.071	-0.065	-0.831***	-0.028*	-0.088***
	[0.000]	[0.107]	[0.254]	[0.000]	[0.040]	[0.000]
DV mean (nonparents $r = 24$)	0.14	-0.12	0.21	1.61	12.55	1.31
Unique individuals	8,936	6,511	8,220	8,936	8,936	8,936
Observations	130,519	80,094	55,866	1,155,300	1,160,802	1,155,300
R^2	0.26	0.27	0.19	0.38	0.76	0.79
			B. Fathers			
1-month effect	-0.124***	-0.038*	0.025	-0.057*	0.007*	0.001
	[0.000]	[0.019]	[0.227]	[0.035]	[0.031]	[0.877]
12-month effect	-0.046***	-0.008	0.052**	-0.034	0.008	-0.003
	[0.000]	[0.592]	[0.004]	[0.343]	[0.075]	[0.618]
24-month effect	-0.008	0.030	0.009	0.036	0.014**	0.012
	[0.461]	[0.076]	[0.665]	[0.427]	[0.007]	[0.085]
DV mean (nonparents $r = 24$)	0.02	-0.08	0.25	2.01	12.84	0.96
Unique individuals	54,726	47,266	45,454	54,726	54,726	54,726
Observations	1,871,046	875,571	669,685	12,659,523	12,654,617	12,659,523
R^2	0.25	0.29	0.18	0.40	0.81	0.76

Notes: Displays predicted values from Eq. 2, the semi-parametric specification. Outcomes include (1) standardized (mean=0, SD=1) scores from physical/combat fitness tests conducted 2x per year, (2) standardized scores (mean=0, SD=1) from supervisor-rated job performance evaluations conducted 1-2x per year, (3) standardized scores (mean=0, SD=1) from rifle or pistol tests conducted 1 or fewer times per year, (4) cumulative months of training, (5) cumulative degree counts relative to $r = -10$, and (6) cumulative promotion counts relative to $r = -10$. We exclude women's physical performance scores 9 months before through 7 months after birth because women are not required to take fitness tests during and after pregnancy. We exclude women's marksmanship scores 9 months before through 4 months after birth because women are not required to take marksmanship exams during pregnancy or while on leave. All outcomes for women and men exclude $r = 0$. Regressions include match-group and month-by-year fixed effects. Predicted p-value of whether the value statistically differs from zero are shown in brackets, based on heteroscedasticity-robust F-test and standard errors clustered by match-group and individual. $^{***}\ p < 0.001$, $^{**}\ p < 0.01$, $^{*}p < 0.05$.

Table III: Identifying Potential Mechanisms for Changes in Promotion

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female X Parent	-0.099***	-0.071***	-0.091***	-0.101***	-0.101***	-0.099***	-0.072***	0.003	0.002
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.017)	(0.017)
Parent	0.018**	0.019**	0.015*	0.018**	0.018**	0.018**	0.017**	0.013*	0.013*
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Female	0.108***	0.106***	0.092***	0.109***	0.108***	0.108***	0.092***	0.118***	0.100***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Avg post physical		0.084***					0.075***		0.067***
		(0.004)					(0.004)		(0.004)
Avg post job			0.127***				0.115***		0.113***
			(0.004)				(0.004)		(0.004)
Avg post mark				-0.012*			-0.020***		-0.021***
				(0.005)			(0.005)		(0.005)
Training (months)					-0.002**		-0.003***		-0.004***
					(0.001)		(0.001)		(0.001)
Education (years)						-0.001	-0.004		-0.004
						(0.002)	(0.002)		(0.002)
# waived fit tests								-0.053***	-0.040***
								(0.003)	(0.003)
Prepregnancy controls	X	X	X	X	X	X	X	X	X
p(fem*par+par)=0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.320	0.357
Sobel-Goodman mediation		0.28	0.08	-0.01	-0.02	-0.00	0.27	1.03	1.02
Observations	158867	158867	158867	158867	158867	158867	158867	158867	158867
R-squared	0.2974	0.3050	0.3101	0.2975	0.2976	0.2974	0.3161	0.3016	0.3184
Adjusted R-squared	0.2974	0.3049	0.3100	0.2975	0.2975	0.2974	0.3160	0.3016	0.3183

Notes: Predicts count of promotions from $r = [-10, 24]$ in a combined women/men placebo sample, using one "wide" observation per match-ID. "Female X Parent" is a mother-specific indicator variable. "Parent" is an indicator for a parent in general. "Prepregnancy controls" includes controls for variables included in he LASSO model (see Appendix Table A.1) plus months of service, an indicator for officer, and an interaction between months of service and officer, and year at $r = -10$. The average post physical, job, and rifle scores are mean observed scores from $r > 0$; we do not drop any observed scores so observations for mothers who choose to take a fitness test soon after birth would be included. Training and education are the total months of training and the total years of education achieved, respectively, by $r = 24$. Number of missed tests is a count of the physical and combat fitness testing periods without an observed fitness test for $r > -10$; this is based on the USMC fitness test schedule rather than relative time to/from birth. We cluster standard errors by ID because some nonparent IDs are matched multiple times to more than one parent. Includes the p-value of an F-test of whether the parent and mother-specific coefficient sum to zero. Sobel-Goodman mediation test computes the proportion of the total mother-specific effect of parenthood on promotion counts that is mediated by the additional sets of controls. $^{***}\ p < 0.001$, $^{**}\ p < 0.01$, $^{*}p < 0.05$.

Table IV: Women's Outcomes by Leave Length

	Post-birth drop		12 months post		24 months post		
	Effect size	*p*	Effect size	*p*	Effect size	*p*	N
			A. Physical performance (sd)				
Main effect:	-0.498***	0.000	-0.293***	0.000	-0.138***	0.000	95,789
Effects by paid leave length:							
6 weeks	-0.448***	0.000	-0.295***	0.000	-0.140**	0.002	95,789
6 weeks + 12 flex	-0.791***	0.000	-0.288**	0.010	-0.379**	0.009	
18 weeks	-0.650***	0.000	-0.282***	0.000	-0.044	0.601	
p(diff), all effects	0.056		0.986		0.126		
			B. Training (months)				
Main effect:	-0.331***	0.000	-0.637***	0.000	-0.771***	0.000	858,694
Effects by paid leave length:							
6 weeks	-0.326***	0.000	-0.623***	0.000	-0.709***	0.000	858,694
6 weeks + 12 flex	-0.405	0.050	-0.474	0.051	-0.529	0.087	
18 weeks	-0.302*	0.017	-0.747***	0.000	-1.181***	0.000	
p(diff), all effects	0.909		0.607		0.055		
			C. Years of education				
Main effect:	-0.018	0.071	-0.032*	0.013	-0.028	0.083	864,759
Effects by paid leave length:							
6 weeks	-0.009	0.448	-0.018	0.226	-0.008	0.656	864,759
6 weeks + 12 flex	-0.060*	0.033	-0.139**	0.008	-0.137**	0.006	
18 weeks	-0.041	0.058	-0.052	0.058	-0.076*	0.017	
p(diff), all effects	0.158		0.059		0.018		
			D. Promotions (#)				
Main effect:	-0.037**	0.003	-0.084***	0.000	-0.090***	0.000	858,694
Effects by paid leave length:							
6 weeks	-0.042**	0.003	-0.079***	0.000	-0.083***	0.000	858,694
6 weeks + 12 flex	-0.047	0.301	-0.171***	0.001	-0.151**	0.005	
18 weeks	-0.009	0.735	-0.071*	0.014	-0.095**	0.007	
p(diff), all effects	0.540		0.165		0.468		

Notes: Regressions only include birth before March 2016 because mothers could not plan for additional leave announced in July 2015 at conception. Outcomes include physical performance, months of training, years of education, and count of promotion. Postbirth drop is measured at 8 months post-birth for physical performance and at 1 month postbirth for all other outcomes. Regressions include match-group and month-by-year fixed effects. The first row replicates the main analysis for the smaller sample. The next rows display a separate regression from the policy interaction model. "6 weeks" is the predicted mother-placebo gap under the 6-week policy (for babies born December 2014 and prior). "6 weeks + 12 flex" is the predicted mother-placebo gap for mothers who gave birth under the 6-week policy but were retroactively given an additional 12 weeks of leave to use before their baby's first birthday after they had returned to work (for babies born January 2015–mid-May 2015). "18 weeks" value the predicted mother-placebo gap for mothers who gave birth knowing they would have 18 weeks of leave, 12 of which could be used flexibly before the baby's first birthday (for babies born mid-May 2015–October 2016). "12 weeks" is the predicted mother-placebo gap for mothers who gave birth knowing they would have 12 weeks of leave to use immediately following birth (for babies born January 2017 and later). The final row presents the *p*-value for an *F*-test of whether mother-placebo gaps are the same across all policy periods. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

Table V: Specification checks for alternative approaches for women

	Physical Performance (sd) (1)	Job Performance (sd) (2)	Marksmanship (sd) (3)	Training (months) (4)	Education (years) (5)	Promotions (count) (6)
			A. F-test (prepregnancy effects=0), p-value			
Standard TWFE	0.352	0.736	0.127	–	0.000	–
Stacked TWFE	0.009	0.669	0.023	0.000	0.134	0.210
Placebo event, No time FE	0.096	0.774	0.256	0.148	0.030	0.683
Placebo event TWFE (preferred)	0.151	0.790	0.222	0.113	0.025	0.617
			B. 12-month effect			
Standard TWFE	-0.186***	-0.078*	-0.087*	–	-0.100***	–
	[0.000]	[0.023]	[0.041]	–	[0.000]	–
Stacked TWFE	-0.290***	-0.083*	-0.123*	-0.775***	-0.023	-0.105***
	[0.000]	[0.045]	[0.013]	[0.000]	[0.073]	[0.000]
Placebo event, No time FE	-0.290***	-0.171***	-0.117*	-0.711***	-0.026*	-0.073***
	[0.000]	[0.000]	[0.017]	[0.000]	[0.013]	[0.000]
Placebo event TWFE (preferred)	-0.290***	-0.181***	-0.132**	-0.712***	-0.027*	-0.074***
	[0.000]	[0.000]	[0.006]	[0.000]	[0.012]	[0.000]
			C. 24-month effect			
Standard TWFE	-0.129***	-0.186***	-0.027	–	-0.130***	–
	[0.000]	[0.000]	[0.604]	–	[0.000]	–
Stacked TWFE	-0.174***	-0.015	-0.018	-0.895***	-0.023	-0.136***
	[0.000]	[0.756]	[0.758]	[0.000]	[0.148]	[0.000]
Placebo event, No time FE	-0.179***	-0.052	-0.060	-0.829***	-0.027*	-0.088***
	[0.000]	[0.246]	[0.298]	[0.000]	[0.042]	[0.000]
Placebo event TWFE (preferred)	-0.183***	-0.071	-0.065	-0.831***	-0.028*	-0.088***
	[0.000]	[0.107]	[0.254]	[0.000]	[0.040]	[0.000]

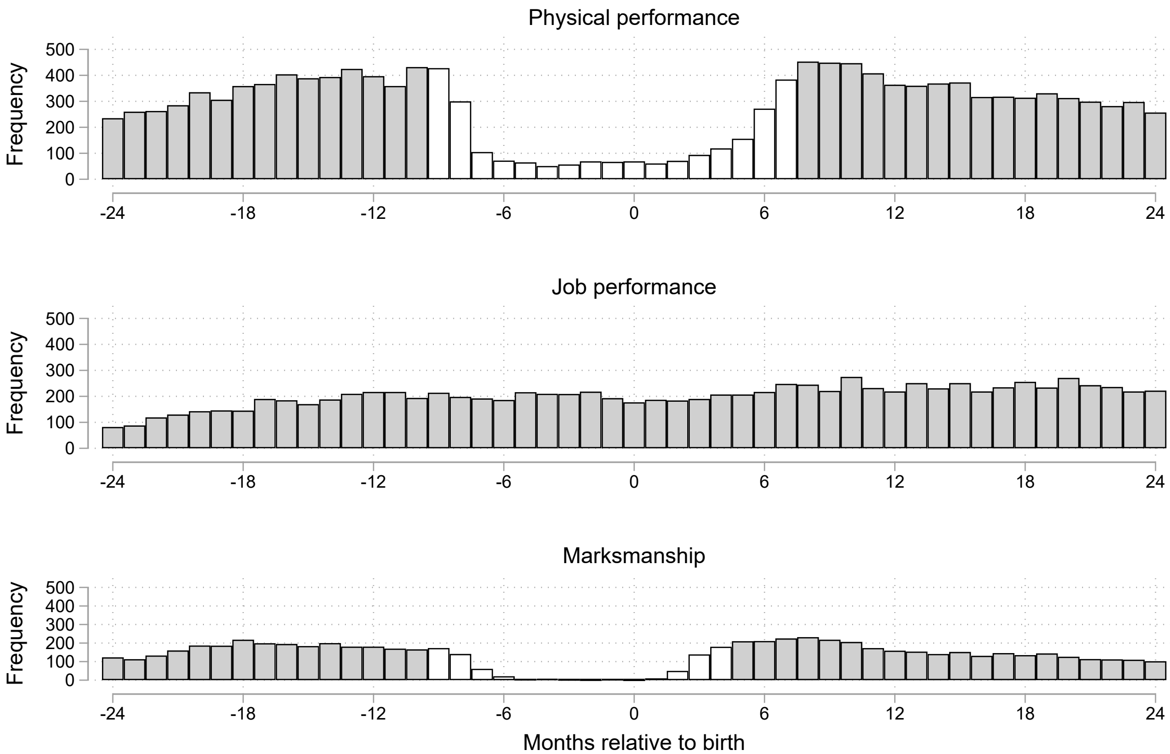
Notes: Tables displays tests and predicted outcomes for alternative specifications for various outcomes. Standard TWFE model is a traditional event study where the comparison group is all same-gender Marines who do not have a baby and remain in the Marine Corps for at least three years; the comparison units are not matched at a particular point in time and do not have estimates for time relative to a placebo birth. Stacked TWFE model uses exact year-month matching with the five nearest neighbors, where time relative to birth and calendar date are synonymous within groups. This model does not exact match on months of service, rank, or active/reserve status; the underlying matching within year-month does include these variables and is identical to the preferred matching model. The model includes estimates for the placebos, then tests whether the parents differ from those patterns. Exact Match, no time FE model is the same matching process as the preferred model, but does only includes $r = [-24, 24]$ rather than binning $r < -24$ and $r > 24$. The model does not include exact month-year fixed effects. The model includes estimates for the placebos, then tests whether the parents differ from those patterns. Exact Match, time FE model is the preferred model that bins $r < -24$ and $r > 24$. The model includes exact month-year fixed effects. The model includes estimates for the placebos, then tests whether the parents differ from those patterns. Details included in Data Appendix. Training and promotion outcomes excluded from the standard TWFE model because it requires a starting point for the count; parents' count starts at $r = -10$. The first panel tests for pretrends with the p-value of an F-test of whether the points estimates for $r = [-24, -11]$ statistically differ from zero. The second and third panel predicts the effect for parents at $r = 12$ and $r = 24$, respectively. *** $p < 0.001$, ** $p < 0.01$, $^{*}p < 0.05$.

Table VI: Specification checks for alternative approaches for men

	Physical Performance (sd) (1)	Job Performance (sd) (2)	Marksmanship (sd) (3)	Training (months) (4)	Education (years) (5)	Promotions (count) (6)
A. F-test (prepregnancy effects=0), p-value						
Standard TWFE	0.804	0.323	0.988	–	0.000	–
Stacked TWFE	0.000	0.949	0.988	0.000	0.000	0.000
Placebo event, No time FE	0.809	0.120	0.465	0.252	0.326	0.399
Placebo event TWFE (preferred)	0.853	0.164	0.354	0.125	0.343	0.274
B. 12-month effect						
Standard TWFE	0.011*	0.059***	0.058***	–	-0.018***	–
	[0.028]	[0.000]	[0.000]	–	[0.000]	–
Stacked TWFE	-0.033***	0.022	0.048***	-0.035	0.015***	-0.043***
	[0.000]	[0.061]	[0.000]	[0.282]	[0.000]	[0.000]
Placebo event, No time FE	-0.045***	-0.003	0.060**	-0.032	0.008	-0.002
	[0.000]	[0.844]	[0.001]	[0.371]	[0.073]	[0.742]
Placebo event TWFE (preferred)	-0.046***	-0.008	0.052**	-0.034	0.008	-0.003
	[0.000]	[0.592]	[0.004]	[0.343]	[0.075]	[0.618]
C. 24-month effect						
Standard TWFE	-0.002	0.007	0.040**	–	-0.018***	–
	[0.759]	[0.538]	[0.004]	–	[0.000]	–
Stacked TWFE	-0.000	0.074***	0.045**	0.057	0.025***	-0.046***
	[0.978]	[0.000]	[0.005]	[0.157]	[0.000]	[0.000]
Placebo event, No time FE	-0.005	0.041*	0.019	0.038	0.015**	0.013
	[0.637]	[0.016]	[0.352]	[0.399]	[0.006]	[0.070]
Placebo event TWFE (preferred)	-0.008	0.030	0.009	0.036	0.014**	0.012
	[0.461]	[0.076]	[0.665]	[0.427]	[0.007]	[0.085]

Notes: Tables displays tests and predicted outcomes for alternative specifications for various outcomes. Standard TWFE model is a traditional event study where the comparison group is all same-gender Marines who do not have a baby and remain in the Marine Corps for at least three years; the comparison units are not matched at a particular point in time and do not have estimates for time relative to a placebo birth. Stacked TWFE model uses exact year-month matching with the five nearest neighbors, where time relative to birth and calendar date are synonymous within groups. This model does not exact match on months of service, rank, or active/reserve status; the underlying matching within year-month does include these variables and is identical to the preferred matching model. The model includes estimates for the placebos, then tests whether the parents differ from those patterns. Exact Match, no time FE model is the same matching process as the preferred model, but does only includes $r = [-24, 24]$ rather than binning $r < -24$ and $r > 24$. The model does not include exact month-year fixed effects. The model includes estimates for the placebos, then tests whether the parents differ from those patterns. Exact Match, time FE model is the preferred model that bins $r < -24$ and $r > 24$. The model includes exact month-year fixed effects. The model includes estimates for the placebos, then tests whether the parents differ from those patterns. Details included in Data Appendix. Training and promotion outcomes excluded from the standard TWFE model because it requires a starting point for the count; parents' count starts at $r = -10$. The first panel tests for pretrends with the p-value of an F-test of whether the points estimates for $r = [-24, -11]$ statistically differ from zero. The second and third panel predicts the effect for parents at $r = 12$ and $r = 24$, respectively. *** $p < 0.001$, ** $p < 0.01$, $^{*}p < 0.05$.

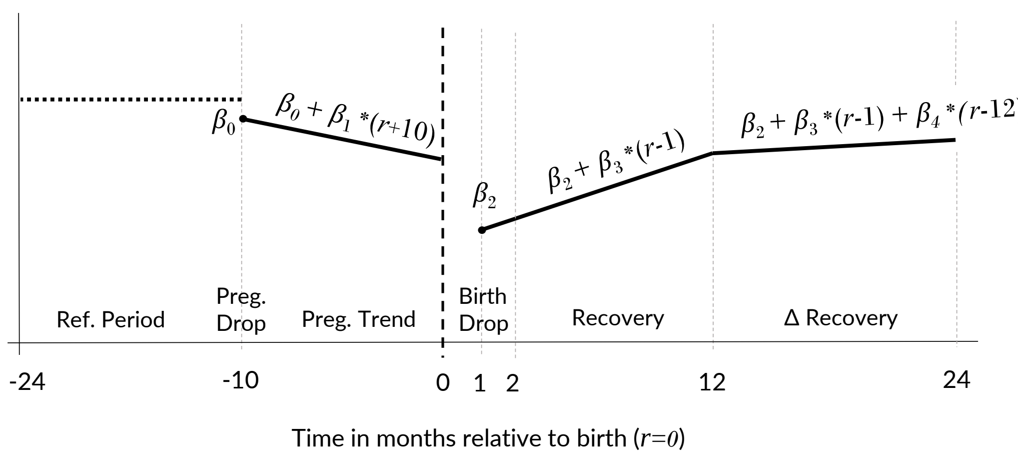
Figure I: Distribution of Mothers' Outcomes Relative to Birth



Notes: Displays the count of physical performance, job performance, and marksmanship scores for mothers by month relative to birth. Physical performance tests are the most common test among these outcomes; all ranks are expected to take them twice a year (the Physical Fitness Test in January–June and the Combat Fitness Test in July–December). Main analysis excludes physical performance scores at $r = [-9, 7]$ (because mothers did not have to take the tests in pregnancy through 6 months postpartum, and commanders may give them some leeway in month 7) and marksmanship scores at $r = [-9, 4]$ (because mothers did not have to take the tests during pregnancy or while on leave). Semi-dynamic specifications always exclude $r = 0$ due to ambiguity about outcome timing relative to birth. Excluded outcomes are in white.

Figure II: Stylized Representation of Parenthood Effects from the Semi-Dynamic Specification (Eq. 2)

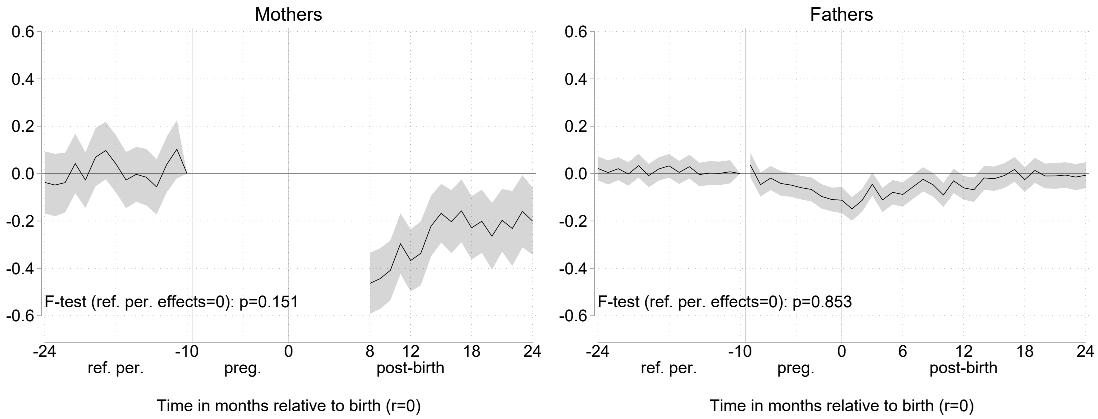
$$Y_{igtr} = \theta_0 PregnancyDrop_{igtr} + \theta_1 PregnancyTrend_{igtr} + \theta_2 BirthDrop_{igtr} + \theta_3 Recovery_{igtr} + \theta_4 \Delta Recovery_{igtr} + \pi P_i + \beta_0 (PregnancyDrop_{igtr} \times P_i) + \beta_1 (PregnancyTrend_{igtr} \times P_i) + \beta_2 (BirthDrop_{igtr} \times P_i) + \beta_3 (Recovery_{igtr} \times P_i) + \beta_4 (\Delta Recovery_{igtr} \times P_i) + \alpha_g + \phi_t + \varepsilon_{igtr}$$



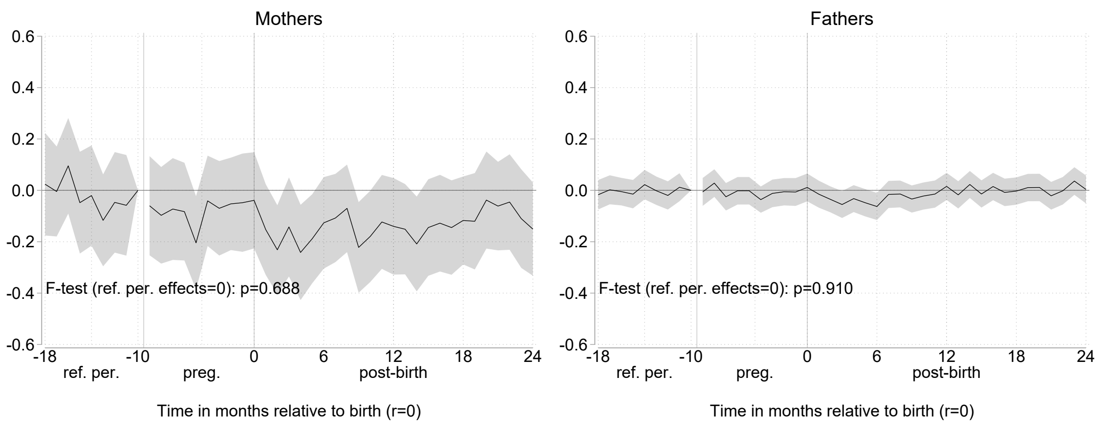
Notes: Figure displays a diagram of parameters defined in Equation 2 where the postbirth drop (β_2) is estimated in the first observed month following pregnancy. For women, we begin measuring the postbirth drop (β_2) for physical fitness performance at 8 months and marksmanship scores at 5 months after the birth. We cannot estimate β_0 or β_1, the pregnancy drop and trend, for women's physical fitness outcomes or marksmanship scores because women are not eligible to be assessed when pregnant.

Figure III: Event Study Estimates of the Impact of Birth on Job Outcomes

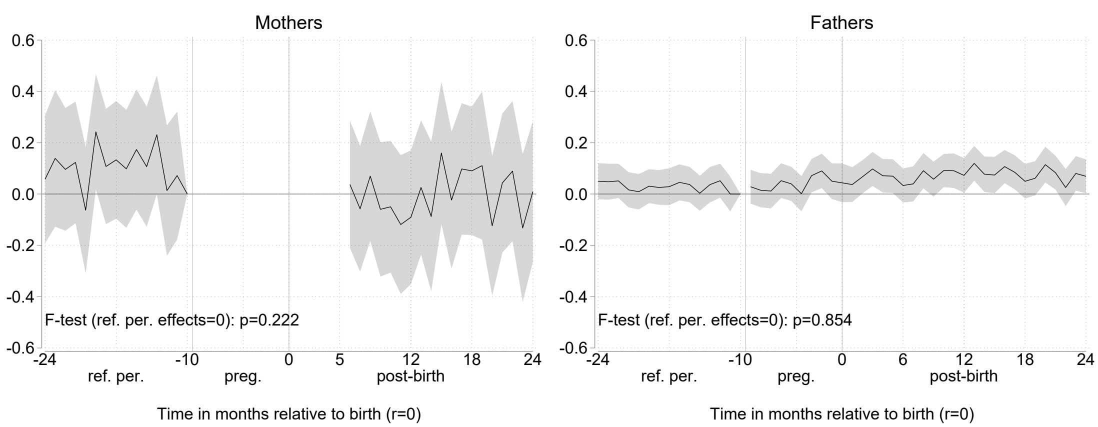
(a) Physical Performance (sd)



(b) Job Performance (sd)



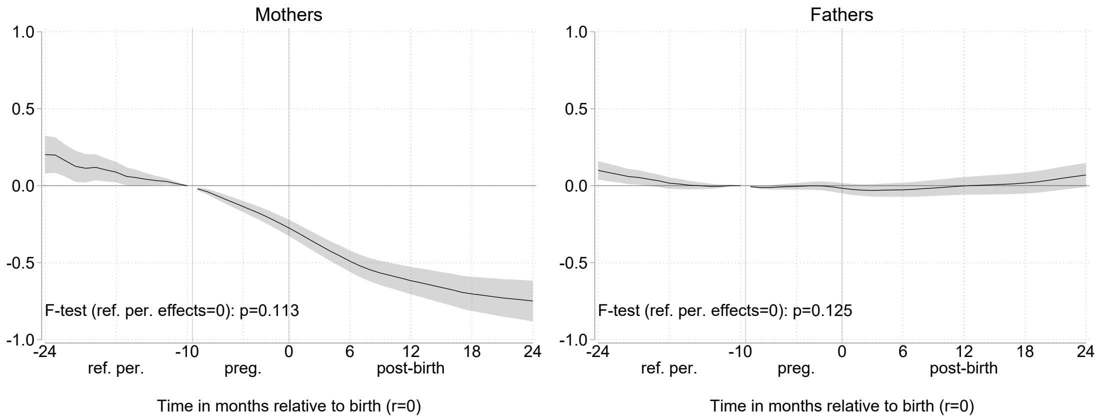
(c) Marksmanship (sd)



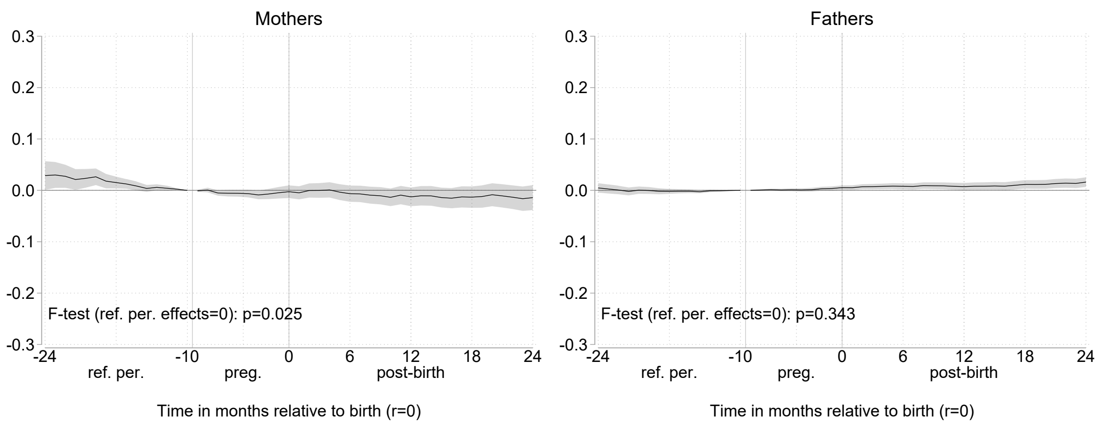
Notes: Displays coefficients from event study regressions using the placebo matched sample. Outcomes include standardized scores from (1) physical/combat fitness tests, (2) job performance evaluations, and (3) rifle/pistol marksmanship evaluations. We require nonparents be an exact match on rank, number of months in service, reserve status, and observation year as parents at $r = -10$. Among those, we match parents to a maximum of five most similar nonparents in their propensity to have a child based on age, race/ethnicity, AFQT scores, marital status (and if their spouse is in the military), education level, months of training, occupational field, and most recent physical performance score as of $r = -10$. Regressions include match-group and month-year fixed effects. The reference month is $r = -10$. Vertical lines reflect the start of the pregnancy ($r = -9.5$) and the birth ($r = 0$). Standard errors are clustered by individual and match-group and are included as shaded areas representing a 95% confidence interval.

Figure IV: Event Study Estimates of the Impact of Birth on Human Capital and Promotions

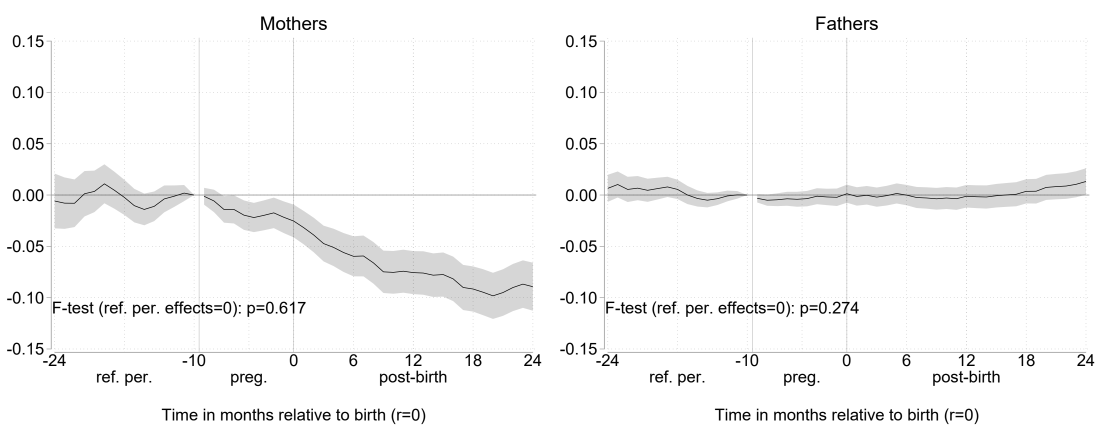
(a) Training (months)



(b) Education (years)



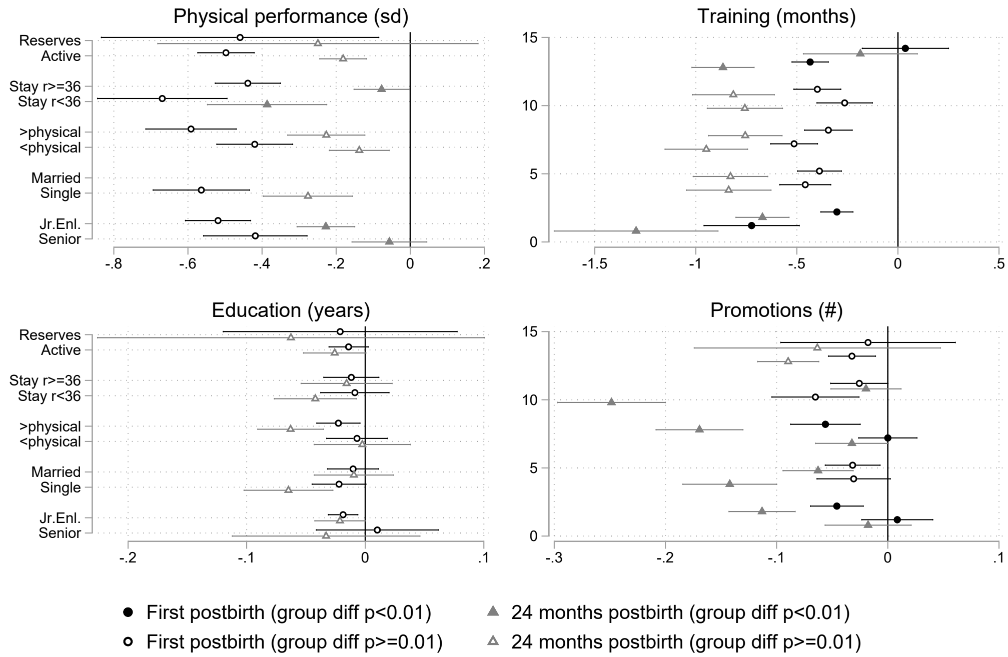
(c) Promotions (#)



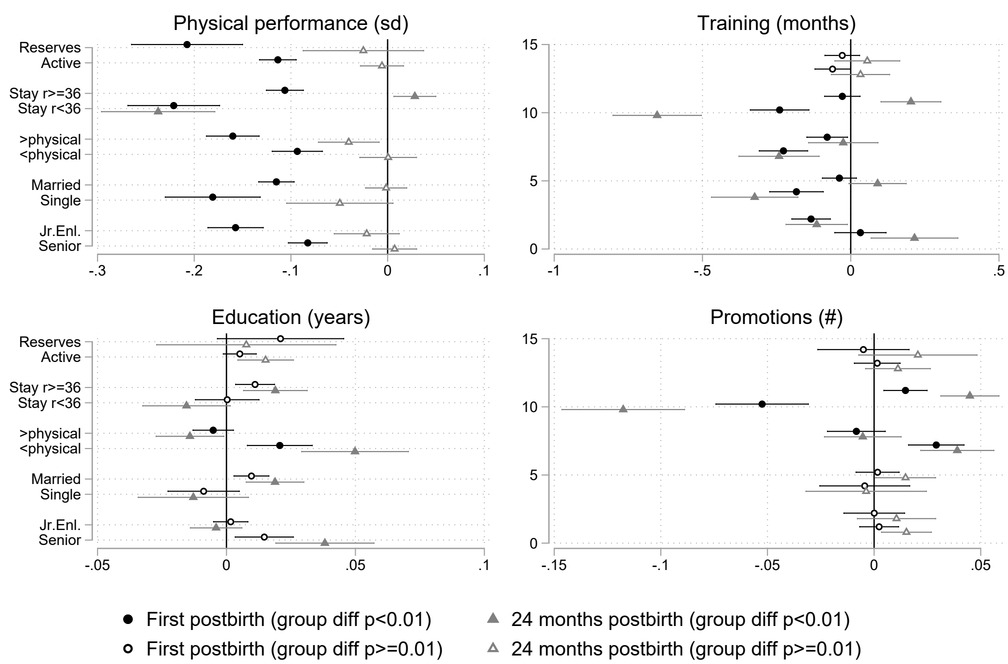
Notes: Displays coefficients from event study regressions using the placebo matched sample. Outcomes include (1) cumulative months of training (relative to $r = -10$), (2) cumulative count of years of education (relative to $r = -10$), and (3) cumulative count of promotions (relative to $r = -10$). We require nonparents be an exact match on rank, number of months in service, reserve status, and observation year as parents at $r = -10$. Among those, we match parents to a maximum of five most similar nonparents in their propensity to have a child based on age, race/ethnicity, AFQT scores, marital status (and if their spouse is in the military), education level, months of training, occupational field, and most recent physical performance score as of $r = -10$. Regressions include match-group and month-year fixed effects. The reference month is $r = -10$. Vertical lines reflect the start of the pregnancy ($r = -9.5$) and the birth ($r = 0$). Standard errors are clustered by individual and match-group and included as shaded areas representing a 95% confidence interval.

Figure V: Event Study Estimates of the Impact of Birth on Job Outcomes

(a) Mothers



(b) Fathers



Notes: Displays gaps in the physical performance, months of training, years of education, and number of promotions relative to prepregnancy between first-time parents and placebo parents across birth events for the first postbirth observation (black line) and 24 month postbirth (light gray line) by subgroups in the placebo sample. Each comparison (e.g., reserve vs. active) is based on one regression by interacting an indicator variable with a group indicator (e.g., reserve) with the parameters in Eq. 2. Classifications for parents are as follows: "Reserves" are not on active duty and likely working a civilian job; "Active" work their military job full-time. "Stay t>36" stay in the military at least 3 years after the birth event; "Stay t<=36" leave between $r = [24, 36]$. ">physical" are those whose military job type above the median physicality level in our sampled based on O*NET classification; "<physical" are at or below the median, among those whose jobs are classified by O*NET. "Married" are married at $r = 0$; "Single" are not. "Jr. Enl" are in enlisted grade E1–E4 at $r = -10$; "Senior" are E5 and up or officers. Vertical solid lines reflect a zero effect. Horizontal lies indicate 95% confidence intervals.

A Supplemental Tables and Figures

Table A.1: Descriptive characteristics for the Placebo Analysis

Variable	Women			Men		
	Mothers (1)	Placebos (2)	Difference (3)	Fathers (4)	Placebos (5)	Difference (6)
A. Exact match variables						
Months of service	39.209	39.209	0.000	58.570	58.570	0.000
	[42.829]	[42.822]	(0.400)	[49.579]	[49.578]	(0.300)
Officer	0.073	0.073	0.000	0.137	0.137	0.000
	[0.260]	[0.260]	(0.003)	[0.343]	[0.343]	(0.003)
Reservist	0.051	0.051	0.000	0.112	0.112	0.000
	[0.221]	[0.221]	(0.002)	[0.316]	[0.315]	(0.002)
B. Other matching variables						
Age	22.565	22.700	-0.134*	24.640	24.907	-0.267***
	[4.172]	[4.345]	(0.059)	[4.623]	[4.989]	(0.037)
Black	0.156	0.150	0.006	0.097	0.081	0.016***
	[0.363]	[0.357]	(0.009)	[0.295]	[0.273]	(0.002)
Hispanic	0.224	0.223	0.001	0.143	0.127	0.017***
	[0.417]	[0.416]	(0.010)	[0.350]	[0.333]	(0.003)
Other	0.097	0.091	0.006	0.073	0.081	-0.008**
	[0.296]	[0.288]	(0.008)	[0.260]	[0.272]	(0.003)
Cognitive test (Z-score)	-0.168	-0.181	0.013	0.022	0.130	-0.108***
	[0.938]	[0.927]	(0.019)	[0.997]	[0.991]	(0.010)
Married	0.413	0.393	0.020**	0.672	0.668	0.004
	[0.493]	[0.488]	(0.007)	[0.470]	[0.471]	(0.004)
Military spouse	0.264	0.251	0.013	0.040	0.037	0.003
	[0.441]	[0.434]	(0.008)	[0.195]	[0.188]	(0.002)
Years of education	12.481	12.487	-0.006	12.743	12.761	-0.018
	[1.328]	[1.324]	(0.022)	[1.557]	[1.576]	(0.013)
Recent fitness score	0.068	0.095	-0.027	0.246	0.195	0.051***
	[0.902]	[0.871]	(0.020)	[0.840]	[0.878]	(0.008)
Combat job type	0.048	0.045	0.004	0.288	0.290	-0.002
	[0.214]	[0.206]	(0.005)	[0.453]	[0.454]	(0.004)
Combat support job type	0.626	0.628	-0.002	0.367	0.352	0.016**
	[0.484]	[0.483]	(0.011)	[0.482]	[0.478]	(0.005)
Aviation job type	0.192	0.199	-0.007	0.242	0.250	-0.008
	[0.394]	[0.399]	(0.009)	[0.428]	[0.433]	(0.005)
Avg. analytic weight	1.000	0.211		1.000	0.202	
Observations	2492	12262	14754	24066	120047	144113
Unique individuals	2492	6444	2492	24066	30660	24066

Notes: Displays means (SD in brackets) for parents (Columns 1 and 4) and their respective placebos (columns 2 and 5), and the difference in means (standard error clustered by person and match group in parentheses) between them (Columns 3 and 6) at the time of the match ($r = 10$), weighted by the analytic weight. Required exact match on months of service, rank (e.g., corporal or captain), reservist, and year, with further matching based on predicted propensity score from the remaining variables and their interactions. Includes the average analytic weight, number of unique person-month matches, and number of unique individuals. *** $p < 0.001$, ** $p < 0.01$, $^{*}p < 0.05$.

Table A.2: Impacts of Childbirth Among First-Time Parents, Controls for Linear Prepregnancy Trends

	Physical Performance (sd)	Job Performance (sd)	Marksmanship (sd)	Training (months)	Education (years)	Promotions (count)
	(1)	(2)	(3)	(4)	(5)	(6)
			A. Mothers			
1-month effect	–	-0.140	–	-0.172**	0.026	-0.033*
		[0.298]		[0.006]	[0.084]	[0.040]
8-month effect	-0.554***	-0.128	-0.096	-0.266**	0.035	-0.060**
	[0.000]	[0.504]	[0.470]	[0.005]	[0.112]	[0.005]
12-month effect	-0.360***	-0.121	-0.107	-0.320**	0.040	-0.075**
	[0.000]	[0.594]	[0.487]	[0.005]	[0.128]	[0.003]
24-month effect	-0.283**	0.017	-0.029	-0.271	0.067	-0.089*
	[0.008]	[0.960]	[0.893]	[0.120]	[0.083]	[0.015]
DV mean (nonparents $r = 24$)	0.14	-0.12	0.21	1.61	12.55	1.31
Unique individuals	8,935	6,510	8,209	8,936	8,936	8,936
Observations	129,875	80,028	55,593	1,150,781	1,156,205	1,150,781
R^2	0.26	0.27	0.19	0.38	0.76	0.79
			B. Fathers			
1-month effect	-0.105***	-0.117*	-0.056	0.070*	0.011*	0.013
	[0.000]	[0.018]	[0.208]	[0.033]	[0.038]	[0.178]
12-month effect	-0.015	-0.145	-0.078	0.175**	0.014	0.019
	[0.590]	[0.084]	[0.240]	[0.003]	[0.143]	[0.204]
24-month effect	0.036	-0.171	-0.176	0.334***	0.023	0.045*
	[0.384]	[0.163]	[0.056]	[0.000]	[0.097]	[0.035]
DV mean (nonparents $r = 24$)	0.02	-0.08	0.25	2.01	12.83	0.96
Unique individuals	54,722	47,261	45,400	54,726	54,726	54,726
Observations	1,853,188	874,039	662,258	12,533,762	12,528,629	12,533,762
R^2	0.25	0.29	0.18	0.40	0.81	0.77

Notes: Displays predicted values from a version of Eq. 2, the semi-parametric specification, that includes a linear slope parameter to control for any prepregnancy trends. All parameters in this model are measured relative to $r = -10$, 10 months before pregnancy, rather than relative to the entire prepregnancy period. See Table II for additional notes. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

Table A.3: Coefficients for Impacts of Childbirth Among First-Time Parents

	Physical Performance (sd)	Job Performance (sd)	Marksmanship (sd)	Training (months)	Education (years)	Promotions (count)
	(1)	(2)	(3)	(4)	(5)	(6)
			A. Mothers			
Pregnancy	–	-0.068	–	-0.061*	-0.015**	0.003
		(0.053)		(0.029)	(0.005)	(0.008)
Pregnancy trend	–	-0.004	–	-0.028***	-0.001	-0.002*
		(0.008)		(0.003)	(0.001)	(0.001)
Postbirth $(birth - 24mos.)$	-0.494***	-0.174***	-0.106	-0.410***	-0.014	-0.032**
	(0.038)	(0.041)	(0.072)	(0.044)	(0.008)	(0.010)
Recovery $(birth - 24mos.)$	0.051***	-0.001	-0.004	-0.027***	-0.001	-0.004***
	(0.012)	(0.004)	(0.013)	(0.003)	(0.001)	(0.001)
Δ recovery $(13 - 24mos.)$	-0.042**	0.010	0.009	0.018***	0.001	0.003
	(0.014)	(0.007)	(0.018)	(0.003)	(0.001)	(0.002)
DV mean (nonparents $r = 24$)	0.14	-0.12	0.21	1.61	12.55	1.31
Unique individuals	8,936	6,511	8,220	8,936	8,936	8,936
Observations	130,519	80,094	55,866	1,155,300	1,160,802	1,155,300
R^2	0.26	0.27	0.19	0.38	0.76	0.79
			B. Fathers			
Pregnancy	0.008	0.020	0.052	-0.035*	-0.001	-0.004
	(0.013)	(0.019)	(0.031)	(0.015)	(0.002)	(0.004)
Pregnancy trend	-0.014***	-0.004	-0.006	0.001	0.000	0.000
	(0.002)	(0.003)	(0.005)	(0.002)	(0.000)	(0.001)
Postbirth $(birth - 24mos.)$	-0.124***	-0.038*	0.025	-0.057*	0.007*	0.001
	(0.009)	(0.016)	(0.021)	(0.027)	(0.003)	(0.005)
Recovery $(birth - 24mos.)$	0.007***	0.003	0.002	0.002	0.000	-0.000
	(0.001)	(0.002)	(0.002)	(0.002)	(0.000)	(0.000)
Δ recovery $(13 - 24mos.)$	-0.004*	0.000	-0.006	0.004*	0.001	0.002*
	(0.002)	(0.003)	(0.004)	(0.002)	(0.000)	(0.001)
DV mean (nonparents $r = 24$)	0.02	-0.08	0.25	2.01	12.84	0.96
Unique individuals	54,726	47,266	45,454	54,726	54,726	54,726
Observations	1,871,046	875,571	669,685	12,659,523	12,654,617	12,659,523
R^2	0.25	0.29	0.18	0.40	0.81	0.76

Notes: Displays coefficients from Eq. 2, the semi-parametric specification. Outcomes include (1) standardized (mean=0, SD=1) scores from physical/combat fitness tests conducted 2x per year, (2) standardized scores (mean=0, SD=1) from supervisor-rated job performance evaluations conducted 1-2x per year, (3) standardized scores (mean=0, SD=1) from rifle or pistol tests conducted 1 or fewer times per year, (4) cumulative months of training, (5) cumulative degree counts relative to $r = -10$, and (6) cumulative promotion counts relative to $r = -10$. We exclude women's physical performance scores 9 months before through 7 months after birth because women are not required to take fitness tests during and after pregnancy. We exclude women's marksmanship scores 9 months before through 4 months after birth because women are not required to take marksmanship exams during pregnancy or while on leave. All outcomes for women and men exclude $r = 0$. Regressions include match-group and month-by-year fixed effects. The parameter "Pregnancy drop" captures any immediate shift from pre-birth to pregnancy, if observed. The parameter "Pregnancy trend" captures trends during pregnancy, if observed. "postbirth drop" is an indicator equal to 1 after the birth, starting in $r = 1$ for all men's outcomes; $r = 8$ for women's physical performance; $r = 5$ for women' marksmanship; and $r = 1$ for women's job performance, training, education, and promotion. "Recovery trend" estimates monthly changes in the outcome for the entire postbirth period. "Δ Recovery trend" estimates any change in the slope in the second year postbirth. Robust standard errors are clustered by match group and individual, shown in parentheses. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

Table A.4: Impact of Childbirth on Physical Performance Test Components

	3-Mile Run (seconds) (1)	Crunches (count) (2)	Pull-Ups (count) (3)	880-Yard-Run (seconds) (4)	Lifts (count) (5)	Shuttle Run (seconds) (6)
			A. Mothers			
8-month effect	55.121***	-5.079***	-1.227**	10.294***	-3.159***	10.824***
	[0.000]	[0.000]	[0.001]	[0.000]	[0.000]	[0.000]
12-month effect	24.368***	-4.010***	-0.696**	7.126***	-2.635***	7.824***
	[0.000]	[0.000]	[0.004]	[0.000]	[0.000]	[0.000]
24-month effect	19.258**	-2.498***	-0.705**	2.065	-1.538*	1.387
	[0.008]	[0.000]	[0.004]	[0.076]	[0.012]	[0.304]
DV mean (nonparents $r = 24$)	1561.69	98.75	7.11	214.67	69.52	186.16
Unique individuals	8,918	8,925	4,730	8,871	8,872	8,871
Observations	70,436	70,778	21,109	69,101	68,830	69,094
R^2	0.23	0.23	0.31	0.19	0.29	0.20
			B. Fathers			
12-month effect	0.595	-0.080	-0.203***	0.700**	0.062	0.887**
	[0.734]	[0.501]	[0.000]	[0.003]	[0.657]	[0.001]
24-month effect	-1.512	0.081	-0.067	-0.301	0.285	0.002
	[0.497]	[0.637]	[0.325]	[0.322]	[0.202]	[0.996]
DV mean (nonparents $r = 24$)	1395.10	102.83	17.16	180.10	94.51	147.83
Unique individuals	54,707	54,716	54,715	54,688	54,467	54,688
Observations	978,269	982,545	969,207	956,306	690,969	956,259
R^2	0.24	0.34	0.23	0.18	0.16	0.20

Notes: Displays coefficients from the semi-parametric specification in Eq. 2 for item-level outcomes by fitness test type. Columns 1–3 show performance on the Physical Fitness Test items, assessed from January to June. Limited pull-up outcome data exist for women prior to 2017, during which time they could do push-ups instead. Columns 3–6 show performance on the Combat Fitness Test items, assessed from July to December. The 880-yard-run (Column 4) captures scores on the Movement to Contact drill, designed to mimic the stresses of running under pressure in battle. Lifts (Column 5) measure the number of times a Marine can lift a 30-pound ammunition can overhead. Shuttle run (Column 6) displays timed performance on a 300-yard shuttle run obstacle, called the Maneuver Under Fire drill, which includes crawls, ammunition resupply, grenade throwing, agility running, and the dragging and carrying of another Marine. Robust standard errors clustered by ID in parentheses. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

Table A.5: Identifying Additional Potential Mechanisms for Changes in Promotion

	(1)	(2)	(3)	(4)	(5)
Female X Parent	-0.099***	-0.087***	-0.023	-0.067***	0.018
	(0.016)	(0.016)	(0.018)	(0.016)	(0.018)
Parent	0.018**	0.023***	0.017**	0.018**	0.019**
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Female	0.108***	0.113***	0.118***	0.109***	0.121***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Deployed post		0.063***			0.059***
		(0.006)			(0.006)
Months limited post			-0.011***		-0.001
			(0.001)		(0.001)
Avg post 1st class fit				0.269***	0.221***
				(0.011)	(0.012)
# waived fit tests					-0.037***
					(0.003)
Prepregnancy controls	X	X	X	X	X
p(fem*par+par)=0	0.000	0.000	0.735	0.001	0.032
Sobel-Goodman mediation		0.12	0.77	0.32	1.18
Observations	158867	158867	158867	158867	158867
R-squared	0.2974	0.2989	0.2990	0.3057	0.3087
Adjusted R-squared	0.2974	0.2988	0.2989	0.3056	0.3086

Notes: Predicts whether an individual will remain until at least 36 months after the match birth event ($r = 36$) in a combined women/men placebo sample, using one "wide" observation per match-ID. "Female X Parent" is a mother-specific indicator variable. "Parent" is an indicator for a parent in general. "Prepregnancy controls" adds controls for variables included in Table 1 civilian job classifications and plus being married to a military spouse, months of service, an interaction between months of service and officer, year, an additional "other non-white" race/ethnicity category, and indicator variables for having a combat, combat support, or aviation-type job (relative to jobs not in these categories) at $r = -10$. Deployed post is an indicator equal to one if an individual was deployed at any point in $r > -10$. Months limited post is an count of the months on limited duty (nondeployable) status in $r > -10$. Average post first-class fitness is the percent of fitness test in the "first-class" category in in $r > 0$. Number of missed tests is a count of the physical and combat fitness testing periods without an observed fitness test for $r > -10$; this is based on the USMC fitness test schedule rather than relative time to/from birth. Count of promotions count in from $r = [-10, 24]$. We cluster standard errors by ID because some nonparent IDs are matched multiple times to more than one parent. Includes the p-value of an F-test of whether the parent and mother-specific coefficient sum to zero. Sobel-Goodman mediation test computes the proportion of the total mother-specific effect of parenthood on promotion counts that is mediated by the additional sets of controls. $^{***}\ p < 0.001$, $^{**}\ p < 0.01$, $^{*}p < 0.05$.

Table A.6: Descriptive characteristics for the Policy Analysis

	Length of Paid Maternity Leave			
	6 wks	6 wks + 12 flex	18 wks	p(diff)
Months of service	38.36	40.17	37.91	0.88
Officer	0.06	0.07	0.11	0.04
Reservist	0.06	0.07	0.04	0.43
Age	22.40	22.77	22.65	0.45
Black	0.16	0.12	0.13	0.17
Hispanic	0.19	0.18	0.24	0.13
Other	0.10	0.12	0.11	0.92
Cognitive test (Z-score)	-0.16	-0.07	-0.18	0.50
Married	0.40	0.40	0.48	0.05
Military spouse	0.26	0.26	0.30	0.45
Years of Education	12.40	12.60	12.65	0.01
Recent fitness score	0.07	0.00	0.01	0.48
Combat job type	0.05	0.02	0.08	0.04
Combat support job type	0.63	0.63	0.59	0.40
Aviation job type	0.19	0.21	0.22	0.42
Observations	1423	121	274	8935

Notes: Displays means for mothers by policy period (columns 1–3) and the p-value of an ANOVA test of whether the values differ across groups (column 4 "p(diff)"). Excludes mothers whose first birth occurred March 2016 or later who could have known about extended leave length at the time of conception. Variables are those used in the matching procedure.

Table A.7: Descriptive characteristics for the Policy Analysis

	Length of Paid Maternity Leave				
	6 wks	6 wks + 12 flex	18 wks	12 wks	p(diff)
Months of service	38.36	40.17	40.97	39.43	0.69
Officer	0.06	0.07	0.10	0.07	0.03
Reservist	0.06	0.07	0.04	0.03	0.04
Age	22.40	22.77	23.06	22.42	0.02
Black	0.16	0.12	0.13	0.17	0.18
Hispanic	0.19	0.18	0.26	0.31	0.00
Other	0.10	0.12	0.10	0.07	0.18
Cognitive test (Z-score)	-0.16	-0.07	-0.15	-0.25	0.18
Married	0.40	0.40	0.48	0.38	0.01
Military spouse	0.26	0.26	0.30	0.24	0.17
Years of Education	12.40	12.60	12.67	12.48	0.00
Recent fitness score	0.07	0.00	0.07	0.07	0.88
Combat job type	0.05	0.02	0.07	0.03	0.05
Combat support job type	0.63	0.63	0.59	0.66	0.24
Aviation job type	0.19	0.21	0.21	0.18	0.58
Observations	1423	121	497	393	2434

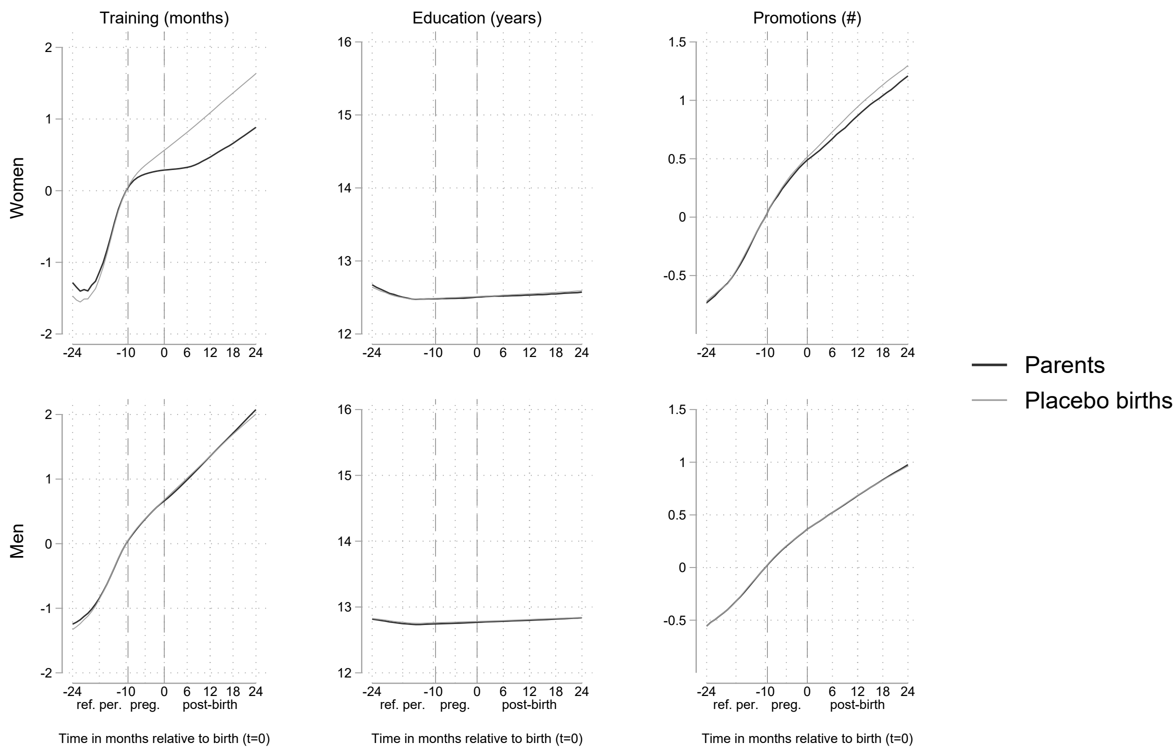
Notes: Displays means for mothers by policy period (columns 1–4) and the p-value of an ANOVA test of whether the values differ across groups (column 5 "p(diff)"). Excludes mothers whose first birth was in November–December 2016 due to ambiguity about the policy for such mothers. Variables are those used in the matching procedure.

Table A.8: Women's Outcomes by Maternity Leave Length

	Post-birth drop		12 months post		24 months post		
	Effect size	p	Effect size	p	Effect size	p	N
			A. Physical performance (sd)				
Main effect:	-0.498***	0.000	-0.293***	0.000	-0.138***	0.000	95,789
Effects by paid leave length:							
6 weeks	-0.448***	0.000	-0.295***	0.000	-0.140**	0.002	95,789
6 weeks + 12 flex	-0.791***	0.000	-0.288**	0.010	-0.379**	0.009	
18 weeks	-0.650***	0.000	-0.282***	0.000	-0.044	0.601	
p(diff), all effects	0.056		0.986		0.126		
			B. Training (months)				
Main effect:	-0.331***	0.000	-0.637***	0.000	-0.771***	0.000	858,694
Effects by paid leave length:							
6 weeks	-0.326***	0.000	-0.623***	0.000	-0.709***	0.000	858,694
6 weeks + 12 flex	-0.405	0.050	-0.474	0.051	-0.529	0.087	
18 weeks	-0.302*	0.017	-0.747***	0.000	-1.181***	0.000	
p(diff), all effects	0.909		0.607		0.055		
			C. Years of education				
Main effect:	-0.018	0.071	-0.032*	0.013	-0.028	0.083	864,759
Effects by paid leave length:							
6 weeks	-0.009	0.448	-0.018	0.226	-0.008	0.656	864,759
6 weeks + 12 flex	-0.060*	0.033	-0.139**	0.008	-0.137**	0.006	
18 weeks	-0.041	0.058	-0.052	0.058	-0.076*	0.017	
p(diff), all effects	0.158		0.059		0.018		
			D. Promotions (#)				
Main effect:	-0.037**	0.003	-0.084***	0.000	-0.090***	0.000	858,694
Effects by paid leave length:							
6 weeks	-0.042**	0.003	-0.079***	0.000	-0.083***	0.000	858,694
6 weeks + 12 flex	-0.047	0.301	-0.171***	0.001	-0.151**	0.005	
18 weeks	-0.009	0.735	-0.071*	0.014	-0.095**	0.007	
p(diff), all effects	0.540		0.165		0.468		

Notes: Regressions exclude birth after March 2016 births given that these women would have known about extended leave before becoming pregnant.Outcomes include physical performance, months of training, years of education, and count of promotion. Postbirth drop is measured at 8 months post-birth for physical performance and at 1 month postbirth for all other outcomes. Regressions include match-group and month-by-year fixed effects. The first row replicates the main analysis for the smaller sample. The next rows display a separate regression from the policy interaction model. "6 weeks" is the predicted mother-placebo gap under the 6-week policy (for babies born December 2014 and prior). "6 weeks + 12 flex" is the predicted mother-placebo gap for mothers who gave birth under the 6-week policy but were retroactively given an additional 12 weeks of leave to use before their baby's first birthday after they had returned to work (for babies born January 2015–mid-May 2015). "18 weeks" values show the predicted mother-placebo difference for mothers who gave birth when 18 weeks of leave was in place but who did not know of this change at the time of conception. For this policy 12 weeks of the leave could be used flexibly before the baby's first birthday (for babies born mid-May 2015–February 2016). The final row presents the p-value for an F-test of whether mother-placebo differences are the same across all policy periods. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

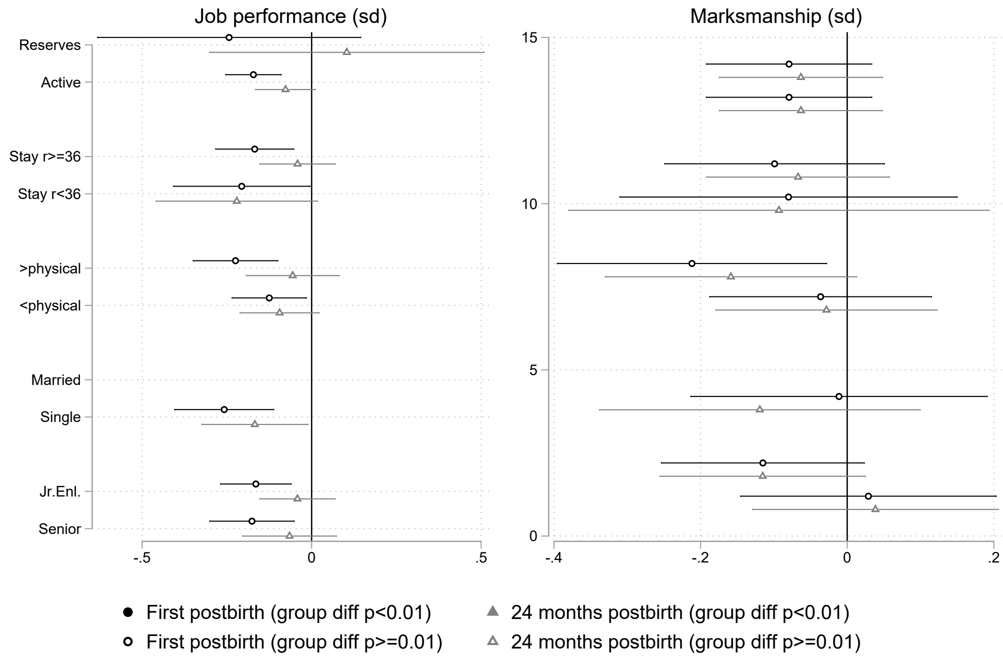
Figure A.6: Placebo Birth Estimates of the Impact of Birth on Human Capital and Promotion



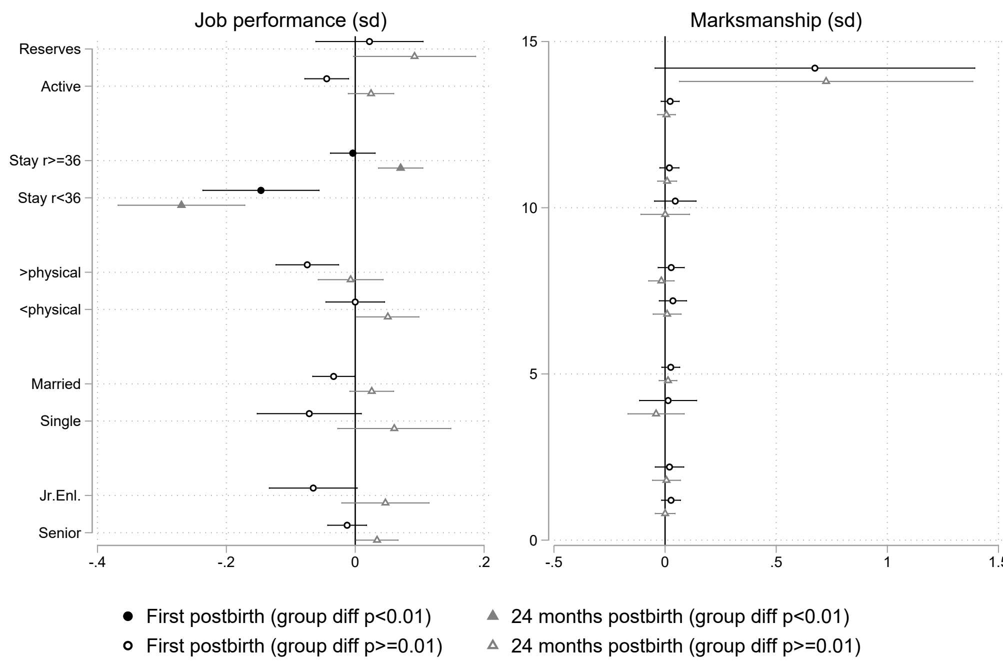
Notes: Displays weighted mean levels of cumulative months of training, cumulative count of years of education (relative to $r = -10$), and cumulative count of promotions (relative to $r = -10$) between first-time parents (solid line) and placebo parents (dashed line) over time. Nonparents assigned to placebo births are limited to those whose rank, number of months in service, reserve status, and year is an exact match with parents' 10 months before the birth. Among those with an exact match, each parent's outcomes are compared to the five nonparents most similar to parents in their propensity to have a child based on age, race/ethnicity, military entrance exam scores (AFQT scores), marital status (including whether a spouse is also in the military), level of education, months of training, occupational field, reserve status, and most recent physical performance scores.

Figure A.7: Event Study Estimates of the Impact of Birth on Job Performance and Marksmanship

(a) Mothers



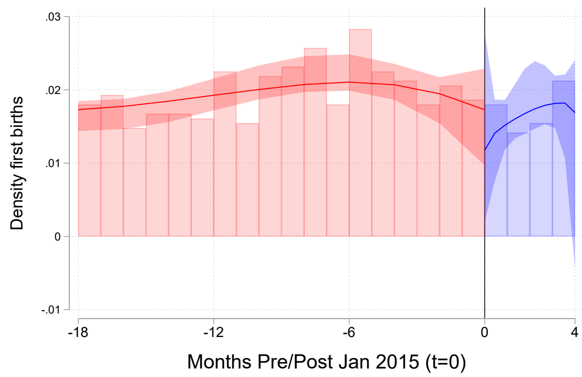
(b) Fathers



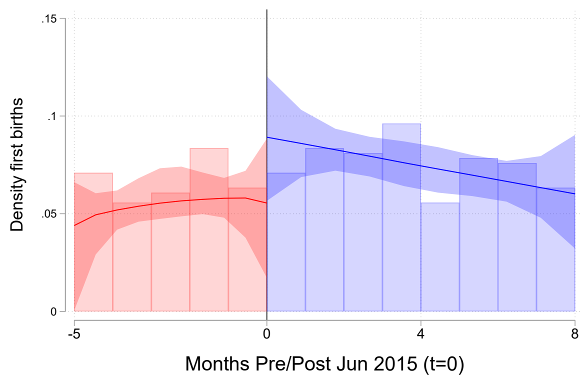
Notes: Displays gaps in the job performance and marksmanship scores relative to prepregnancy between first-time parents and placebo parents across birth events for the first postbirth observation (black line) and 24 month postbirth (light gray line) by subgroups in the placebo sample. Each comparison (e.g., reserve vs. active) is based on one regression by interacting an indicator variable with a group indicator (e.g., reserve) with the parameters in Eq. 2. Classifications for parents are as follows: “Reserves” are not on active duty and likely working a civilian job; “Active” work their military job full-time. “Stay t>=36” stay in the military at least 3 years after the birth event; “Stay t<36” leave between $r = [24, 36]$. “>physical” are those whose military job type above the median physicality level in our sampled based on O*NET classification; “<physical” are at or below the median, among those whose jobs are classified by O*NET. “Married” are married at $r = 0$; “Single” are not. “Jr. Enl” are in enlisted grade E1–E4 at $r = -10$; “Senior” are E5 and up or officers. Vertical solid lines reflect a zero effect. Horizontal lies indicate 95% confidence intervals.

Figure A.8: Density of First Births Across Policy Periods

(a) 6 weeks vs. 6 + 12 flex weeks



(b) 6 + 12 flex weeks vs. 18 weeks



Notes: Histogram bars display the density of first births by month before and after t=0, which differentiates births subject to one leave-length policy period from another. Plotted curves and corresponding 95% confidence intervals come from a manipulation test using a local-polynomial density estimator developed by Cattaneo et al. (2018). The test for a discontinuity at t=0 is not statistically significant in Panel (a) or (b). The sample includes all women in the Marines with a first birth during the time window.